

NPRDC TR 82-56

AUGUST 1982

AD-A119128

EXPECTANCY THEORY MODELING

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED



NAVY PERSONNEL RESEARCH
AND
DEVELOPMENT CENTER
San Diego, California 92152



EXPECTANCY THEORY MODELING

Paul Horst

Reviewed by
Richard C. Sorenson

Released by
James F. Kelly, Jr.
Commanding Officer

Navy Personnel Research and Development Center
San Diego, California 92152

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The history, terms, and concepts of expectancy theory are examined. The basic concepts of multivariate analysis models are discussed and applied to develop a more adequate expectancy theory model. Criticisms of present models and their postulates and assumptions are addressed. Measurement problems are solved by the development of analytical models that will accept data more easily obtained from subjects. Recommendations are made regarding empirical tests of the models.

S/N 0102- LF- 014- 6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

FOREWORD

Work on this report was stimulated by limitations of expectancy theory formulations as they were encountered and employed in exploratory development work unit 521-001-018-03.02; Expectancy Theory of Work Motivation (PE 62/63N), which was funded during the period FY76-79. In connection with that work unit, the author began a thorough reformulation of expectancy theory in terms that are objective and concepts that are measurable. This report is the result of that reformulation. It was prepared in its present form as part of FY82 in-house independent laboratory research work unit ZR000-01-042-04-01.10; Models and Measures of Human Performance (PE 61152N).

This report is intended for use by those conducting research in work motivation and organization management.

JAMES F. KELLY, JR.
Commanding Officer

JAMES J. REGAN
Technical Director

SUMMARY

Problem

Although a vast amount of literature in organizational behavior has been generated by expectancy theory since 1964, this literature has not been substantially influenced by the traditional models of multivariate analysis. Further development and application of expectancy theory requires a better methodological and mathematical foundation than is currently provided.

Objective

An objective of this effort was to reformulate expectancy theory in organizational behavior in objective terms and measurable concepts, employing sound multivariate models.

Approach

The history, terms, and concepts of expectancy theory are examined. The basic concepts of multivariate analysis models are discussed and applied to develop a more adequate expectancy theory model. Criticisms of present models and their postulates and assumptions are addressed.

Conclusions

Expectancy theory offers a particularly good challenge to apply the best taxonomic and mathematical modeling. It addresses the highly important area of motivation that has been well supplied with literary, semantic, and rhetorical effort but lacks an adequate taxonomic and mathematical infrastructure. Analytic models are described that provide a basis for theoretical development and empirical validation.

Recommendation

It is suggested that models be developed to solve for scale and origin parameters, relieving the subject of making these judgments. Several candidate designs are provided.

CONTENTS

| | Page |
|--|------|
| INTRODUCTION | 1 |
| History | 1 |
| Expectancy Theory and Traditional Models | 1 |
| Terms Used in Expectancy Theory | 2 |
| Motivation | 2 |
| Effort | 2 |
| Performance | 2 |
| Outcome | 3 |
| Need | 3 |
| Basic Concepts of Expectancy Theory | 3 |
| Valence | 3 |
| Expectancy | 4 |
| Instrumentality | 7 |
| VROOM MODELS | 9 |
| The Concept of Choice | 9 |
| The Models | 9 |
| The Valence Model | 9 |
| The Behavior Choice Model | 9 |
| Suggested Notational Changes | 10 |
| Modification of the Vroom Models | 11 |
| Need for Two Distinct Models | 11 |
| Notational Changes | 11 |
| BASIC CONCEPTS IN MULTIVARIATE ANALYSIS MODELS | 14 |
| The Simple Data Matrix and its Terminology | 14 |
| The 3-Mode Data Matrix | 14 |
| The 3-Modality Set | 14 |
| Differing Terminology | 15 |
| Levels | 16 |
| Confusion in the Use of Terms | 16 |
| Levels as a Dimensional Group | 16 |
| Outcomes | 17 |
| Determination of the Number of Levels | 18 |
| Measurement Scales | 18 |
| DIMENSIONS OF A MODALITY | 20 |
| Levels | 20 |
| Manifest vs. Latent Dimensions | 20 |
| Predictor vs. Criterion Dimensions | 21 |
| Variables | 21 |
| Scales | 21 |
| Predictor and Criterion Variables | 21 |
| The Interaction Dimension | 22 |
| Definition | 22 |
| Moderator Variables | 22 |
| Generalized Models | 22 |

| | Page |
|---|------|
| Important Role of Interaction. | 22 |
| Concerns Over the Use of Interaction Variables | 23 |
| Linear vs. Nonlinear Relationships. | 23 |
| Intervention Variables | 23 |
| The Importance of Intervention Variables. | 23 |
| The Special Case of Predictor Variables | 24 |
| N-ary Predictor Variables | 24 |
| Practical Suggestions for Intervention Variables. | 24 |
| Experimentation-utilization | 25 |
| Endogenous-exogenous Predictor Dimensions | 25 |
| RELATIONSHIPS AMONG VARIABLES | 26 |
| The Simple Variable Function. | 26 |
| The Two-parameter Model | 28 |
| The Loss Function. | 29 |
| The Constraints. | 30 |
| Solving for the Parameters | 30 |
| Summary for the Two-parameter Case | 32 |
| The Canonical Correlation Model | 32 |
| The Parameters. | 33 |
| The Loss Function. | 33 |
| The Constraints. | 34 |
| The Solution | 35 |
| Significance Tests. | 35 |
| The General Multivariate Case | 35 |
| The Mathematical Model. | 36 |
| The Loss Function. | 36 |
| The Constraints. | 36 |
| The Solution | 36 |
| Significance Tests. | 37 |
| Special Cases of Significance Tests | 38 |
| Kinds of Relationships | 38 |
| Linear Relationships. | 38 |
| Nonlinear Relationships | 38 |
| Summary of Multivariate Analysis Systems | 39 |
| The Structural Component | 39 |
| The Functional Component | 39 |
| Optimal Parameter Estimation | 40 |
| Examples of Multivariate Modeling | 40 |
| A MORE ADEQUATE EXPECTANCY THEORY MODEL | 40 |
| The Environments. | 40 |
| The Organism | 41 |
| The Interactive vs. the Expectancy System | 42 |
| The Organismic Modality. | 42 |
| The Attribute Modality | 42 |
| The Job Environment Modality | 42 |
| The Focal Environment Modality | 42 |
| The Contingency Environment Modality | 43 |
| The Restructured Focal Environment Modality | 43 |

| | Page |
|--|------|
| The Outcome Modality | 43 |
| The Needs Modality | 44 |
| Summary of Modalities | 44 |
| Summary of the Interactive System | 45 |
| Valence and Expectancy | 45 |
| Valence in Terms of Alternatives | 45 |
| Concomitance of Valence and Expectancy | 46 |
| Expectancy and the Operant Event | 46 |
| Expectancy and Experience. | 47 |
| Expectancy and Certainty | 47 |
| EXPECTANCY THEORY AS A MULTIVARIATE ANALYSIS SYSTEM. | 48 |
| The Abstract Model | 48 |
| A Notational System | 49 |
| The Instrumentality Concept | 50 |
| The Modality of Instrumentality. | 50 |
| Instrumentality as a Correlation | 50 |
| The Levels of Dimension | 52 |
| Two-level Dimensions | 52 |
| Multi-level Dimensions | 52 |
| Expectancies for Pairs of Dimensions | 53 |
| The Nebeker-Moy Model | 53 |
| The Algebra of the Model | 53 |
| Revision of the Nebeker-Moy Model | 56 |
| Possible Sources of Confusion. | 56 |
| The Algebra of the Model | 56 |
| Kinds of Expectancies | 58 |
| Can and Will Expectancies | 58 |
| Special Nebeker Models | 59 |
| THE UNRESTRICTED LEVEL MODEL | 59 |
| The Two-dimensional Case | 59 |
| The Mathematical Notation | 59 |
| The Correlation Between Two Outcomes | 60 |
| Expectancies and the Correlation Model | 61 |
| Instrumentality with Unrestricted Levels. | 62 |
| The Multidimensional Case | 63 |
| Mathematical Development | 63 |
| Practical Considerations. | 65 |
| The Criterion Modality | 65 |
| REVIEW OF THE EXPECTANCY THEORY MODEL | 65 |
| The Structural Elements | 65 |
| The Performance Modality | 67 |
| The Outcome Modality. | 67 |
| The Needs Modality | 68 |
| The Valence Modality | 68 |
| The Expectancy Modality | 68 |

| | Page |
|---|------|
| DIMENSIONS OF MODALITIES | 68 |
| Levels of Dimensions | 69 |
| Homogeneity of Dimensions | 69 |
| Specifications of Dimensions | 70 |
| The Comprehensiveness of Dimensions | 70 |
| The Specificity of Dimensions | 71 |
| Dimensions and Verbalizations | 71 |
| Latent vs. Manifest Dimensions | 72 |
| Factor Analytic Methods | 72 |
| Limitations of Factor Analysis | 73 |
| Dimensions and the Needs Modality | 74 |
| Summary | 75 |
| REVISION OF VROOM'S MODEL | 75 |
| The Mathematical Development. | 75 |
| Definition of Notation | 75 |
| The Mathematical Formulation | 75 |
| Introduction of the Parameters | 76 |
| Extension to the Person Modality | 77 |
| The Matrix Formulation | 78 |
| The Least Square Solution | 79 |
| Implications of the Generalized Solution | 80 |
| A Concrete Example | 81 |
| Expectancy Theory and the Multivariate Analysis Model | 81 |
| The Within-subject vs. Across-subject Analysis | 82 |
| The Vector-to-Matrix Transition | 82 |
| Limitation of the Within vs. Across Distinction | 83 |
| Reducing the Number of Parameters. | 83 |
| LEVELS OF DIMENSIONS | 83 |
| The Job Choice Model | 84 |
| The Performance Level Model | 84 |
| Summary of Criterion Types | 84 |
| The Contingency Modality | 85 |
| Units and Origin of Assigned Values | 85 |
| The Valence Modality | 86 |
| SCALE ORIGIN AND INTERACTION | 86 |
| The Schmidt Critique | 86 |
| The Multiple Regression Alternative. | 86 |
| The Values of Levels | 87 |
| The Number of Levels | 87 |
| The Mathematics of the Scale and Origin Problem | 87 |
| Definition of Notation | 87 |
| The Mathematical Development. | 88 |
| The Case of Unit Weights | 90 |
| Constant Transformation Parameters | 91 |

| | Page |
|--|------|
| Variable Transformation Parameters. | 92 |
| Modeling for Scale and Origin Parameters | 93 |
| Linearity or Nonlinearity. | 94 |
| Quantification of Concepts. | 94 |
| THE THREE-MODALITY MODEL | 94 |
| The Galbraith-Cummings Model. | 95 |
| Modification of the Model | 95 |
| Definition of Notation. | 96 |
| The Scalar Algebra Formulation. | 96 |
| The Matrix Algebra Formulation | 97 |
| The General Case | 98 |
| INTERVENTION VARIABLES | 99 |
| The Contingency Variables | 99 |
| The Expectancy Variable. | 100 |
| Summary of Intervention Variables | 100 |
| THE POSTULATES OF BEHLING AND STARKE. | 100 |
| Introduction to the Postulates | 100 |
| The Postulates | 102 |
| Comparability | 102 |
| Transitivity | 103 |
| Independence. | 107 |
| Dominance. | 109 |
| Optimization. | 110 |
| Reconstitution | 111 |
| Equivalent n-Fold Partition. | 112 |
| Irrelevance of Identical Outcomes. | 113 |
| Ambiguity | 113 |
| THE RATIONALITY ASSUMPTIONS OF WAHBA AND HOUSE | 113 |
| Comparison of Rationality Assumptions with the Postulates. | 114 |
| Continuity Assumption. | 114 |
| The Statement of the Assumption | 114 |
| Critique of the Statement | 115 |
| Mathematical Formulations of the Assumption | 116 |
| Weaknesses of the Continuity Statement | 118 |
| CONCLUSIONS. | 119 |
| REFERENCES | 121 |

| | Page |
|--|------|
| APPENDIX A--OPTIMAL ROW TRANSFORMATIONS OF PAIRED MATRICES | A-0 |
| APPENDIX B--OPTIMAL ROW TRANSFORMATIONS OF PAIRED MATRICES: THE SINGLE VECTOR CASE | B-0 |
| APPENDIX C--A SIMPLIFIED MODEL FOR OPTIMAL ROW TRANSFORMATIONS OF PAIRED MATRICES | C-0 |
| APPENDIX D--SIMILARITY TRANSFORMATIONS FOR PAIRED MATRICES. | D-0 |
| APPENDIX E--THE WEIGHTED AVERAGE ROW SUM AND VARIANCE MODEL | E-0 |
| APPENDIX F--THE PAIRED MATRIX MODEL WITH EQUAL ROW AND COLUMN AVERAGE VARIANCES. | F-0 |
| APPENDIX G--MINIMUM VARIANCE TRANSFORMATION VECTORS FOR PAIRED MATRICES. | G-0 |
| APPENDIX H--MAXIMUM CORRELATION TRANSFORMATION VECTORS FOR PAIRED MATRICES WITH EQUAL COLUMN MEANS AND VARIANCES | H-0 |
| APPENDIX I--THE NEBEKER MODEL | I-0 |
| APPENDIX J--A PROOF OF THE VALIDITY OF A NEBEKER EXPECTANCY THEORY MODEL | J-0 |

INTRODUCTION

History

Expectancy theory in organizational behavior has generated a vast amount of experimentation and literature in the comparatively brief span of 15 years. While much of the inspiration for this great volume of effort goes back many years, it is only since the work of Vroom (1964) that the major portion of the work has been published. Mitchell's excellent review (1974) lists over 60 articles and books that appeared in the 10 years following Vroom's (1964) book on the subject of expectancy theory in work situations. This number has been increasing steadily each year. The chapter by Campbell and Pritchard in the Handbook of Industrial and Organizational Psychology (1976) is devoted almost exclusively to expectancy theory in organizational behavior, including a consideration of its origins and of alternative theories. It lists upwards of 120 references. This extensive preoccupation of some psychologists is all the more remarkable considering that, outside a limited circle, few psychologists are aware of the great activity that has taken place in the application of expectancy theory to organizational behavior in recent years.

Expectancy Theory and Traditional Models

Interestingly enough, a perusal of the literature to date leads to the conclusion that the activity in the field that has taken place has not been influenced by the traditional models of multivariate analysis. This lack is particularly striking because even the mainstream of multivariate theory appears neither broad nor deep enough to accommodate the phenomena and concepts explicit or implied by the expectancy models. As we shall attempt to show, the modeling implications of expectancy theory require a more rigorous methodological and mathematical underpinning than is currently provided by available multivariate models and a far greater attention to these considerations than has heretofore been accorded to them.

It is interesting to note that typical introductions to expectancy theory point out the simplicity of the model. However, further pursuit of the literature reveals a bewildering plethora of semantic confusion, ambiguities, and the use of imprecisely defined concepts. It is as though the theory builders were undertaking the construction of an enormously complicated structure without benefit of plans or specifications.

Nevertheless, as one explores and ponders the vast semantic jungle of expectancy theory literature, intriguing outlines begin to emerge that suggest how plans and specifications might be drawn up. Based on these outlines, a well structured and integrated comprehensive model of expectancy theory might be constructed from which interesting and perhaps useful special cases could be developed.

This approach is admittedly at odds with the proposal of Campbell and Pritchard (1976) to the effect that we should abandon the search for an overall grand design and concentrate in depth on fragmented bits and pieces from the salvage pile of narrow psychological theories or fads. I believe this latter approach characterizes most of the work psychologists have been engaged in over the last 40 years and that this accounts for the fact that the psychological profession has had and continues to have such limited impact on the ongoing activities of the educational, industrial, governmental, and other major components of our social system.

Terms Used in Expectancy Theory

In any case, expectancy theory has been applied to a variety of behavior phenomena in spite of the fact that it has not yet succeeded in formulating an overall grand design. On the assumption that it may prove even more generally useful in the prediction and control of human behavior if a somewhat more ambitious design than is currently available should be developed, we shall proceed to delineate the outlines of a more comprehensive and general model.

One attractive feature of expectancy theory, as developed to date, is its adherence to the emphasis of a few basic concepts, such as expectancy and valence. However, before these are discussed, we shall discuss several terms that recur in the literature. The first of these is motivation which, although used frequently, plays a limited role. The other terms we will discuss are effort, performance, outcome, and needs.

Motivation

Most discussions of expectancy theory, together with efforts at model building, appear to have major interest in the concept of motivation. As a matter of fact, the rather definitive chapter of Campbell and Pritchard (1976), devoted mainly to expectancy theory in organizational behavior, emphasizes motivation in the title and early on introduces the subtitle, "What is motivation?" The question is never clearly answered throughout the rather lengthy chapter and perhaps it is just as well that no serious attempt is made to answer it. In any case, it usually turns out that, in the literature we have seen, no one seriously comes to grips with any strictly operational definition of the word. Perhaps this is all by way of suggesting that the term is of doubtful value and need not be regarded as a useful word in a comprehensive expectancy model.

Effort

Effort presumably is an act or response of a person or organism, directed toward the achievement of a task-goal. The discussions of effort are typically at a less than highly sophisticated level but one can accept the term without too much questioning for the present at least.

Performance

Performance does not yield so readily to common sense understanding. Whether it is something that the subject is or has been doing in the achievement of a task-goal or whether it is the achieved task-goal itself is not always clear. Whether it is the Cheshire cat, its grin, or both cannot always be known for certain. In any case, it is something associated with the task-goal accomplishment. However, this raises at once a major difficulty encountered in the literature of expectancy theory. One can often not be sure when two or more different words or phrases refer to the same thing, or when the same term is used to refer to several or more different concepts or phenomena. Not that this problem is restricted to expectancy theory. It permeates much of the literature of psychology in general as well as that of other social sciences. If we seem to be beating a dead horse, it is only by way of emphasizing the importance of unambiguous terminology in the formulation of a tightly structured complex model.

One point that the literature does and should emphasize is that effort and performance are not the same thing. Effort precedes performance but factors other than effort, such as ability, may influence performance.

Outcome

Outcomes may follow either effort or performance or both. However, not everything that follows effort or performance is an outcome. It is an outcome only if it may have value for those who exert effort or perform. The term valence is sometimes defined as anticipated value, which presumably means that valence is the value that a person thinks an outcome will have at the time it occurs. This distinction between value and valence may turn out to be important in the evolution of a comprehensive expectancy model but so far it does not appear to have played a crucial role in expectancy theory models. Perhaps a space for such a distinction should be left in the table of atomic elements of an expectancy theory model, just in case.

Need

Need satisfactions may follow outcomes and indeed these latter may acquire their valence from the fact that they lead to need satisfaction. This statement raises an epistemological issue that we shall not belabor here. We merely suggest that the commonly used phrase "leads to" carries the same occult overtones as other words (e.g., "explain," "cause," "effect," "understand"), that plague semi- and pseudoscientific literature. However, for want of a less theistically tinged expression and to avoid involved circumlocutions, we shall employ the phrase "leads to" for the time being.

Sometimes need satisfaction is referred to as a second level outcome. Presumably, needs have come to take the place of instincts, drives, etc., but their current status is not clearly defined and will require further discussion later in the text. Occasionally the literature has given lip service to the need for needs in discussions of the expectancy models but little if any use of the term appears to have been made in experimental studies.

It is also curious that no attempt has been made to put the concept on the same basis as effort, performance, and outcomes. This concept seems to have more importance in the overall model than it has been accorded and perhaps a term more nearly coordinate with the other sets of events should be sought or coined.

Basic Concepts of Expectancy Theory

Valence

Until a pressing need arises to distinguish further between "valence" and "value," we use the term valence to refer to the amount of value that an event is expected to have for a person. This raises the question of whether the traditional designation of outcome in the chain effort-performance-outcome-need is well chosen, but we shall consider this question at greater length later. For now, we shall merely say that any of the four sets of events may have valence for a person. Exertion of effort may have valence, as may performance of a task-goal, an outcome, or a need satisfaction. We may have positive, zero, or negative valence for any event or thing. We shall not try at this time to distinguish between valence and such other valuing concepts as "desirable," "satisfying," "important," etc., and their antonyms. This is not to deny that a need for such distinction may develop. For now, we shall follow the practice of including such valuing adjectives in the overall term valence.

Presumably, this term valence implies a most fundamental type of avoidance-approach or removal-sustaining response with reference to a stimulus pattern and gives it firm status with reference to core concepts of leading learning and behavior theorists.

For this reason, among others, the concept may be trusted to play a crucial role in formulating a more definitive theoretical model.

We see then that the concept of valence is fairly clearly defined, that it is applicable to a wide variety of events, and that it appears to be of fundamental psychological importance. We have also suggested a number of kinds of events to which the concept of valence may be applied.

Expectancy

We shall consider for the present only briefly the concept from which the theory under discussion takes its name: expectancy.

Types of Expectancy. As used in the literature, expectancy refers to a person's subjective belief with respect to the occurrence of an event. We shall see that there may be at least two kinds of expectancy with reference to the occurrence of an event. One of these is the expectation of a person that the event will occur and the other is his expectation that he can effect the occurrence. This second type of expectation has not been emphasized in the literature but it has been introduced into the experimental design of some expectancy theory research (Dockstader, Nebeker, Nocella, & Shumate, 1980). A detailed analysis of the distinction between the two types of expectancy might suggest how both could be profitably integrated into a comprehensive model.

Behavior and Behavior Modification. The concept of expectancy may appear unduly abstract for a tightly structured theory of human behavior but it should not be too difficult to relate it to more operational and traditional concepts of behavior in general and behavior modification in particular.

The Expectancy Concept and Learning Theory. It is a little surprising that the use of the expectancy concept has come largely from the literature on decision theory, with its narrow philosophical basis in abstract probability theory, rather than in some of the more enduring concepts and principles of learning and behavior modification in general. Perhaps this anomaly is due to the fact that each psychologist has tended to align himself or herself with a single one of the leading learning theorists, such as Thorndike, Hull, Lewin, Tolman, Skinner, to say nothing of the more recent formulators of mathematical models. It is true, of course, that Hull used the concept in the early thirties.

There has been a tendency to emphasize and magnify the differences among the various theories rather than to distill the common core of all in terms of more primitive and fundamental conceptualizations. This tendency may arise, in part, from either the boredom or disenchantment of psychologists with the earlier models based on simple stimulus-response concepts. I suspect that the answer is to be found more in boredom than in disenchantment. Granted that phenomena could be observed that would not yield readily to the more naive Pavlovian models, one may still suspect that the temptation to seek refuge in rhetorical and semantic solutions may have won over the application of more rigorous methodological and even mechanistic approaches.

In the earlier days, when psychologists were not thought to be naive if they couched their theories in stimulus-response terminology, they did not hesitate to talk about the probability of a given response following a given stimulus. The learning of a given task consisted of increasing this probability. Thorndike's law of effect was invoked as an explanatory principle but he did not seem to find it necessary to emphasize the role of stimulus-response associations in the further explication of the law. The notion of rewards and punishments for correct and incorrect responses was invoked to account for

increasing the probability of the correct response. Some theorists even went so far as to define reward and punishment as stimuli that evoked, respectively, approach and avoidance responses. A few emphasized that an approach response was one that tended to maintain the "rewarding" stimulus and an avoidance response was one that tended to separate the subject and the punishing stimulus but, for the most part, this interpretation was accepted explicitly or implicitly.

The Descriptive Chain. It was not so clearly spelled out, however, that the entire descriptive chain called not only for the probability that a given stimulus will be followed by a given response but for other probabilities as well. In the first place, the organism has a repertoire of responses, each one of which could follow the given stimulus. To each of these, we may assign a probability that it will follow the given stimulus. There exists also a set of stimuli, each of which may be characterized by the extent to which it evokes a response, that eliminates or maintains the response-evoking property of the stimulus. This characteristic of the stimulus is what we have described as the valence of an outcome or, more generally, an event. It appears to be what the expectancy theorists refer to as the valence of an outcome. If the stimulus is followed by a stimulus-extinguishing response, it is negatively valent and if it is followed by a stimulus-sustaining response, it is positively valent.

We may now speak of the probability that, for a specified stimulus, some response in the organism's repertoire will occur. We may also speak of the probability that, for some responses, a stimulus of specified valence will occur. Other things being equal, we assume that the more frequently a response following a given stimulus is followed by a positively valent stimulus, the greater the probability that the given stimulus will be followed by that response. Likewise, the more frequently a response following a given stimulus is followed by a negatively valent stimulus, the lower the probability that the given stimulus will be followed by that response.

The Expectancy Construct in the Chain. Certainly none of this is new; the literature is replete with elaborations that consider variations of delayed reinforcement, reinforcement schedules, and secondary reinforcements, to say nothing of all the changes that have been rung on a vast array of conditioning and conditioners. However, we do need to see just where the expectancy construct fits into this oversimplified stimulus-probability-response-probability-valent stimulus chain.

First we note there has been much agonizing, perhaps less rather than more fruitful, of the precise mechanisms, neural or otherwise, by which the probability in the stimulus-response association becomes modified by probability of the succeeding stimulus and its valence. Many years ago, Dashiell (1949) suggested some interesting learning theories involving neural networks and the modification of synaptic resistance. Unfortunately, as is frequently the case, too many were quick to point out deficiencies in the theory and too few were willing or able to undertake the harder task of improving the theory. Thus, nothing much has ever come of it. Perhaps it is just as well. This type of speculation leads more often than not into the fruitless infinite regression inquiry into what causes the causes that cause effects.

Let us return now to the question of where the expectancy construct fits into our chain of stimuli, responses, probability, and valence. It should not be too difficult to guess that the notion of a subject's estimate of the probability that a given response will be followed by a given valent stimulus comes close to the way in which the term has been used in expectancy theory models. However, we need to examine briefly at this point what we mean by the subject's estimate of the probability that an event will occur. This

examination is necessary because we wish to exorcise as early and completely as possible any excess philosophical connotations of terms and constructs.

It can be assumed that a person's expectancy that an event will occur is based on his past experience, directly or indirectly through observation or communication with others, that, under comparable conditions, it has occurred a given proportion of the times. This is not a rigorous or all-inclusive definition but perhaps it is sufficiently operational to dispel most anxieties that might be associated with more occult implications of the term. More directly, we choose a historical or experiential interpretation of the term rather than some more threatening philosophical or mathematical explanation.

We have gone to some length to establish the status of the term "expectancy" in our model-building enterprise, since it has played such an important role in the current literature and since, in all probability, it will be called upon to carry even a heavier burden in a more comprehensive expectancy theory model. However, before suggesting how this increased responsibility may be thrown onto the term, we shall discuss several issues related to the probability concept in stimulus-response mechanism models.

One of these issues arises because of the nature of the valent stimuli. Actually there are at least two important kinds of valent stimuli. One of these might be called an organismic stimulus or an intrinsically valent stimulus and the other, a manipulable or extrinsically valent stimulus. The intrinsically valent stimulus is administered by the organism to itself. The subject exerts effort toward the accomplishment of a task goal and the mere exertion of effort may give him pleasure or pain, satisfaction or dissatisfaction. As a matter of fact, the effort or organismic activity may not be directed toward the accomplishment of any prespecified task goal. The organism may get satisfaction or discomfort from merely random flexing or tensing of body musculature or from idle or random meditation, fantasizing, or what not. Also, he may experience satisfaction or dissatisfaction from the contemplation of a task goal for which he has achieved a more or less adequate degree of completion. The point is that, in either case, whether restricted to effort, directed or random, or whether extended to include a self-appraisal of a task goal achievement, the rewards or punishments are applied by the organism to itself.

The second kind of valent stimulus is one over whose occurrence the organism has little if any direct control. It is typically initiated by an external agency. In the work situation, this may be supervisory or policy-making personnel who provide salary increases, promotions, commendations, or so on. On the negative side they may include reprimands, penalties, or actual dismissal. These types of valent stimulus situations may also be administered in the form of expressions of approbation or disapproval by fellow workers or others in or outside the actual working environment.

Whether the valent stimulus situations are intrinsic or extrinsic, the organism presumably will have some expectancy that they will follow some specified activity on his part. Before we can discuss in greater detail the implications of the two types of expectancies, we will need to develop the characteristics of the structural elements of a more adequate model in greater detail. For the present, we merely emphasize that the expectancy of external valent stimuli is subject to modification by organizational management and as such may have important implications for the achievement of overall mission goals of the organization.

Reward Versus Punishment. Before proceeding to a discussion of an expanded role of the expectancy concept in organizational behavior, we shall comment briefly on a second issue related to the probability concept in stimulus response mechanism models. The

question has been repeatedly raised over the years as to which is more effective, reward for good behavior or punishment for bad. Apparently the question has never been satisfactorily answered, nor should it be, for like many questions it has rarely been satisfactorily asked. The early analytical work on learning theory by Thurstone (1930) and by Gulliksen (1934) included the parameters of positive and negative reinforcement in the development of rational learning curves. This work led into the more recent quantitative approach to learning theory that has come to be designated by the rather ambitious and pretentiously inclusive designation of mathematical models. However, for the most part, the mathematical formulations that have emerged appear to be minimally concerned with variation in the algebraic signs of reinforcement type parameters.

It appears to be well accepted, however, that the introduction of negatively valent stimuli for inappropriate responses that occur at a low probability level can be highly disruptive of behavior and counterproductive for learning. It is also currently popular within certain ideological and political circles to regard the administration of any but positively valent stimuli as inappropriate for behavior modification in the rehabilitation of socially deviant members of society. It is highly probable that the domesticators and trainers of subhuman animals learned many centuries ago by trial and error the optimal mix of positively and negatively valent stimuli for attaining prespecified goals of behavior modification throughout the training periods of their animal wards. In any case, we shall venture a set of hypotheses that, although presently only tangentially related to a comprehensive expectancy theory model, may eventually have promise for incorporation into a further expanded model. The hypotheses are:

1. The lower the probability of a correct response, the higher should be the reward of a correct response and punishment should be high only for highly probable incorrect responses.
2. The higher the probability of a correct response, the greater should be the punishment for all incorrect responses and the lower the reward for correct ones.

These hypotheses suggest a shifting from reward for correct responses to punishment for wrong responses as the probability of correct responses increases, with a maintenance of punishment for all highly probable incorrect responses at all levels of correct response probability.

Instrumentality

We suggest then that the concepts of valence and expectancy, when divested of extra-scientific implications, may become, at the same time, simple but important concepts in the construction of a comprehensive human behavior model. We have indicated that they have indeed assumed a key role in the development of expectancy theory in organizational behavior to date.

We shall now introduce a third concept that has assumed a major role in expectancy theory. This is instrumentality. In fact, some writers use the acronym VIE (valence, instrumentality, expectancy) theory rather than the term expectancy theory. It is interesting to note that some writers go to great length to describe the role and meaning of instrumentality in expectancy theory, while others question the utility of the concept and indeed recommend that it be dispensed with as such.

Differences from Expectancy. The term instrumentality was introduced by Vroom (1964). He and others concede that, while it is closely related to expectancy, it is

different in at least two ways. One of these has to do with the way the concept is quantified and the other with the type of events with which it is associated.

Taking up first the quantification aspect, we recall that valence can be either positive or negative. While the treatment of the quantification of valence has not been notable for precision and clarity, there appears to be general agreement that the limits to which valence may vary, both positively and negatively, are not necessarily restricted. This permissiveness in the specification of range may well underlie some of the unresolved issues with which expectancy theory continues to struggle.

However, there is general agreement that expectancy shall be treated as a traditional probability that, by definition, can vary only between zero and plus one. Instrumentality, on the other hand, is treated as a correlation coefficient and therefore is restricted to a range of minus one to plus one. In some models experimenters have encountered difficulties or inconsistencies when attempting to combine mathematical valences and instrumentalities associated with events. The difficulties have arisen due to the admission of both positive and negative values for the two variables, and some have questioned the logic of allowing both positive and negative values for instrumentalities. Although it is probable that a more sophisticated mathematical formulation might avoid the logical difficulties, there are more fundamental objections to the concept of instrumentality itself that will emerge as our discussion continues.

A second distinction that is made between expectancy and instrumentality has to do with the type of events with which it is associated. It says that expectancy is an act-to-outcome relationship while instrumentality is an outcome-to-outcome relationship. This distinction introduces several questions. First, why is it necessary to posit a different type of relationship for the two cases? Why cannot one have an expectancy that one outcome will lead to another outcome, just as well as an expectancy that one act will lead to another? Certainly a probability is a much simpler concept mathematically than a correlation coefficient. Some have even gone so far as to suggest two types of expectancy: one as the probability of goal accomplishment, given a particular individual situation, and the other the probability of receiving a first-level outcome, given achievement of the task goal. Finally, the first level outcomes (rewards) are conceived as having instrumentality for second-level outcomes (need reduction). Another schema for the sequence of relationships posits (1) an expectancy that a specified level of effort will or will not accomplish the task, (2) the instrumentality of task performance for job outcomes, and (3) the instrumentality of outcomes for need satisfaction.

We see then that investigators differ in the types of events to which expectancies and instrumentalities are assigned. In the two examples cited above, both types of relationships are assigned to the performance-outcome sequence of events.

One of the problems with attempting to distinguish between expectancy and instrumentality on the basis of the type of events that they purport to relate is the lack of clarity exhibited in the definitions of these events. An outcome has been defined as anything a person might wish to attain. But then one might wish to attain a high level of effort or a high level of task goal accomplishment, both of which might be regarded as acts or behaviors rather than outcomes to which they might lead.

Questionable Usefulness of Instrumentality Concept. In any case, it is becoming clear that neither the probability versus correlation nor the act-outcome versus outcome-outcome are consistent or useful bases for distinguishing between expectancy and instrumentality. Since expectancy is the simpler of the two concepts and since we may be able to show that we can get along at least as well, if not better, without the concept of

instrumentality, we may suggest here that this concept of instrumentality be dispensed with and leave for subsequent discussion a more in-depth analysis for justifying its liquidation.

VROOM MODELS

The Concept of Choice

It is necessary at this point to augment the four-category set of events that were introduced early in this discussion: effort, performance, outcomes, and needs. In order to present briefly the conventional approach to the Vroom models, another concept is required; namely, choice. Originally Vroom suggested two models, an important merit of both being that they may be expressed as mathematical equations, which, at least, gives them the appearance of precision.

The Models

The Valence Model

The first of the Vroom models has been called the valence model and may be written

$$V_j = f \sum_{k=1}^n (V_k I_{jk}) \quad (1)$$

where:

V_j = the valence of outcome j ,

I_{jk} = the cognized instrumentality of outcome j for the attainment of outcome k ,

V_k = the valence of outcome k , and

n = the number of outcomes.

The equation is interpreted to mean that V_j is a monotonically increasing function of the summation of products on the right side of the equation. Although it is not always stated explicitly, outcome j is assumed to be temporally antecedent to the group of outcomes that includes the n outcomes corresponding to the n V_k 's. Serious problems are involved in the metrics of the V 's; these and other problems will be taken up later.

Two important features of the valence model are that it includes (1) valences of outcomes only and (2) instrumentalities but not expectancies.

The Behavior Choice Model

Vroom's second model has been called the behavior choice model. It may be written

$$F_i = \sum_{j=1}^n (E_{ij} V_j) \quad (2)$$

where:

- F_i = the force on the individual to perform act i,
- E_{ij} = the strength of the expectancy that act i will be followed by outcome j,
- V_j = the valence of outcome j, and
- n = the number of outcomes.

This equation is also significant in several respects. First, it includes no instrumentalities but, rather, expectancies, E_{ij} , that outcome j will follow act i. Second, it includes a variable, F_i , the force on an individual to perform act i. This variable appears to be a sort of shadow or phantom variable associated with act i rather than some distinct attribute of act i.

The notion of a force or tendency to act could be as difficult for the hard-nosed experimentalist to grapple with as an other-world apparition. Is this F_i any more or less than a valence estimate of act i?

Suggested Notational Changes

To foreshadow later discussion, it is suggested that, in the first equation, V_j is not really the valence of outcome j but only an estimate of it and that, instead of writing V_j in the equation, it might be better to write \tilde{V}_j , where the tilde means "estimate of." Since we have shown our strong inclination to dispense with the instrumentality concept and to include it in the expectancy concept, we suggest replacing I with E in the equation. Then the first equation could be written

$$\tilde{V}_j = f \left(\sum_{k=1}^n (E_{jk} V_k) \right) \quad (3)$$

where, without loss of generality, we have reversed the order of the two variables on the right.

Next let us look again at equation (2). Suppose we follow our suggestion that, without losing essential information, we let F_i be the estimated valence of act i and indicate it by the symbol \tilde{V}_i . Suppose also that we recognize that equation (2), as written, cannot be taken too seriously but must be generalized with an f preceding the symbols on the right to indicate "a monotonically increasing function of." We may then write

$$\tilde{V}_i = f \left(\sum_{j=1}^n (E_{ij} V_j) \right). \quad (4)$$

A comparison of equations (3) and (4) reveals that they are the same except for the seemingly small matter of subscripts. There are also some minor details that we shall clear up. First, the symbol f precedes the symbols on the right of both equations. This would imply that the functions for both are the same, which would not necessarily be the

case. We shall therefore introduce subscripts for the f of equation (4) to give f_{ij} and for the f of equation (3) to give f_{jk} . We also note that the upper limit on the summation sign in both equations is n , which implies that the number of outcomes in the second set of events for equation (3) is the same as the number of outcomes for the second set of events in equation (4). There is no justification in general for this assumption; therefore, we substitute m for n in the upper limit of the summation in equation (4). We now write equations (4) and (3), respectively, as:

$$\tilde{V}_i = f_{ij} \left(\sum_{j=1}^m (E_{ij} V_j) \right) \quad (5)$$

and

$$\tilde{V}_j = f_{jk} \left(\sum_{k=1}^n (E_{jk} V_k) \right). \quad (6)$$

In passing, we call attention to some additional symbols that we have sneaked into these equations. For each one, we have interposed an opening parenthesis on the right side between the f and the remaining symbols. We have also, for both, added a closing parenthesis at the end of the equation. These two inclusive pairs of parentheses mean that the functional symbol applies to the entire notational set within them.

We see now that the two equations (5) and (6) imply three chronologically ordered sets of events. Equation (5) implies on the left a set of events whose typical member is i , followed by a second set of events whose typical member is j . The second equation picks up with this second set of events and indicates that it is followed by a third set of events whose typical member is k .

Modification of the Vroom Models

Need for Two Distinct Models

We may now enumerate a few points suggested by this discussion of the two Vroom models, the valence model and the job choice model. First, the need for two distinct models may be seriously questioned. Do we really gain anything by considering them as separate models rather than as special chronological cases of a more generally ordered chronological model? It may be argued that the act-outcome sequence is essentially different from the outcome-outcome sequence. However, if we dispense with the instrumentality concept and replace it with expectancy, we imply that acts and outcomes are not essentially different types of events that call for modeling essentially different from that called for by prior and subsequent events in general. Since, as will be seen later, the model can become highly complex in spite of all we can do to restrain it, we shall strive to eliminate distinctions for which there is no readily demonstrable need.

Notational Changes

A second point that the notational discussion suggests is equally important. It may appear that the apparent compulsiveness with the notational changes is little more than nitpicking. Skilled verbalists could argue that only a sketchy symbolic presentation, such as is characteristic of the literature on expectancy theory, is enough to give the presentation the appearance of scientific and mathematical sophistication. They could

insist that the crucial concepts and relationships can be adequately, if not better, explicated by well-wrought phrases. We suggest that this is far from the truth. Even the more sophisticated of multivariate analysis models cannot be adequately developed by the most skillful of verbalizations. Only the parsimony and precision of a carefully wrought notational structure is adequate for unambiguous formulations of these models. Even mathematicians who should know better are too often content with inefficient and ambiguous symbolic systems that require extensive and diffuse verbal elaboration for adequate clarification. However, as suggested earlier, a more adequate model of expectancy theory can well extend beyond the frontiers of currently existing multivariate analysis models in complexity and demand an even more rigorous and meticulous attention to an efficient notational system. This system cannot be thrown together in bits and pieces as afterthoughts but must be carefully developed as the evolution of a more comprehensive theory proceeds.

At the risk of being tedious, but without apology, we further pursue the notational issue by considering Vroom's general force model. This is given as

$$F_i = \sum_{j=1}^n E_{ij} \left(\sum_{k=1}^n I_{jk} V_k \right) \quad (7)$$

where

$$\begin{aligned} F_i &= \text{force toward act } i, \\ E_{ij} &= \text{expectancy that act } i \text{ will result in performance } j, \\ I_{jk} &= \text{the instrumentality of performance } j \text{ for outcome } k, \text{ and} \\ V_k &= \text{the valence of outcome } k. \end{aligned}$$

In addition, although the notation is not defined, we must assume that

n is the number of performance events with typical performance j and also the number of outcomes with typical outcome k .

We now, as in equation (4), substitute \tilde{V}_i for F_i . As in equation (3), we substitute E_{ijk} for I_{jk} . Furthermore, we let the number of distinct performance events be m to distinguish it from the number of outcome events n . Then we also introduce another pair of parentheses and rewrite equation (7) as

$$\tilde{V}_i = \sum_{j=1}^m E_{ij} \left(\sum_{k=1}^n (E_{jk} V_k) \right) \quad (8)$$

We note that the expression in the inner parentheses on the right side of (8) is precisely the same as the expression in parentheses on the right side of (6). We shall write from equation (6):

$$\sum_{k=1}^n (E_{jk} V_k) = F_{jk}(V_j). \quad (9)$$

Now the F_{jk} on the right of (9) means that we have taken the inverse function of f_{jk} in (6). We thus remove the f_{jk} from the summation term and apply its inverse function to the \tilde{V}_j term. If any variable y is a monotonically increasing function of another variable x , then it can readily be shown that x is also a monotonically increasing function of y .

Now let us go back to equations (5) and (6). Suppose first we drop the tilde over the \tilde{V}_j in equation (6) and write simply

$$V_j = f_{jk} \left(\sum_{k=1}^n (E_{jk} V_k) \right). \quad (10)$$

For the present, we do not attempt to justify dropping the tilde and later we may find that we had no right to do so. Such a discovery may also lead us to conclude that current formulations may involve doing things that we have no right to do. In any case, suppose now we substitute for V_j in equation (5) the right side of equation (10). This gives us

$$\tilde{V}_i = f_{ij} \left(\sum_{j=1}^m E_{ij} \left(f_{jk} \left(\sum_{k=1}^n (E_{jk} V_k) \right) \right) \right). \quad (11)$$

Suppose again we casually drop the f_{ij} and f_{jk} from equation (11), even though later we may find we had no right to do so. Then the equation becomes

$$\tilde{V}_i = \sum_{j=1}^m E_{ij} \left(\sum_{k=1}^n (E_{jk} V_k) \right). \quad (12)$$

We now see that equation (12) is the same as equation (8) and we recall that this latter equation is a modification of Vroom's general force model given by equation (7).

In these various steps, we have taken some liberties that we have failed to justify. We have failed to distinguish between the estimate of the value of a variable and its observed value. This is an important distinction that is generally recognized in the traditional multivariate analysis models. The symbol with the tilde indicates an estimate and the symbol without the tilde indicates observed values. We shall have more to say later about the meaning of these two terms, but it is probably already obvious that the distinction has to do with dependent and independent variables.

It is bad enough to be fuzzy in our treatment of observed and estimated values, but still more serious to be cavalier in our handling of functional concepts. In our casual introduction and dismissal of the f symbols, we have reflected one of the most serious weaknesses of much of the current work in expectancy theory. The concept of functional relationships between variables is one of the most important in scientific models. The concept involves variables and constants. In a scientific model, the constants assume the role of parameters to be determined and they play a major part in the specifications of a scientific model. Expectancy theory models to date have taken little cognizance of the importance of parameters in the functional specification. It is surprising indeed that expectancy theory in organizational behavior has achieved so much progress in spite of this deficiency. Perhaps this is because of the fundamental soundness of its overall approach to the analysis of human behavior.

BASIC CONCEPTS IN MULTIVARIATE ANALYSIS MODELS

Before we can proceed in our attempts to suggest a more adequate and comprehensive model for expectancy theory in organizational behavior, we must examine in some detail some of the basic concepts in multivariate analysis models in general. Then we may see how expectancy theory concepts may be collated with these concepts or how these latter may be modified or extended to accommodate a comprehensive expectancy theory model.

The Simple Data Matrix and Its Terminology

We shall begin with the simple notion of a data matrix, which is a table of numbers with rows and columns. Each row may represent a person and each column, some attribute or characteristic of the person. For example, the columns may represent different psychological tests and at the intersection of a row and column would be the score that a given person made on that test.

As we proceed, we shall introduce terminology, some of which is already well established and some which is not. Where the terminology is not traditional, we shall attempt to be logical in our choice, or at least to justify our choice if it does not appear to be logical.

First, we note in the example just given that we have two types of things. We have persons and we have attributes or characteristics. We shall call this pair of things a set of things. Each type of thing we shall call a modality. Thus, we have a person modality and an attribute modality. Therefore, we have a set of modalities.

Now the person modality consists of a number of persons and the attribute modality consists of a number of attributes. We shall call each person in the person modality a dimension of the person modality and each attribute in the attribute modality a dimension of the attribute modality. We shall say then that a modality consists of dimensions. In particular, a modality may consist of only one dimension.

We have indicated in the example of the data matrix that, for each person and each test, a score or number is assigned. For each test (attribute), there is a total number of possible values that a test score may take. Each of the possible scores is called a level of the attribute. Thus, for each dimension of the attribute modality there is an ordered series of levels. In the case of the person-attribute set of modalities, we may say that each of the dimensions of the attribute modality has levels. The data matrix consisting of dimensions of a set of person and attribute modalities has, for each distinct pair of person and attribute dimensions, one and only one level of the corresponding attribute dimension.

The question naturally arises as to whether we may just as logically say that, for each dimension of the person modality, there is an ordered series of levels. While conceptually this may seem logical, it will be convenient to consider provisionally two types of modalities—those whose dimensions have levels and those whose dimensions are not always regarded as having levels.

The 3-Mode Data Matrix

The 3-Modality Set

We have considered the example of the simple data matrix. It consists of a set of two modalities. Each of the modalities has dimensions. Each dimension of the attribute

modality has levels. A level of each dimension of an attribute modality is allocated to each dimension of the person modality. (Each person has a score on each test.)

Now the persons may have taken the tests at the beginning of the first semester of the school year. They may also take the tests at intervals throughout the school year. Thus, we have a series of data matrices on the same persons and tests. Since the tests were taken at different times, the persons and conditions may be different so that we may refer to each data matrix as a dimension of a state modality. The totality of these data matrices may be called a three-modality set. For each distinct combination of a dimension from a person, attribute, and state modality, we have a level allocated from the corresponding attribute dimension.

We note that, from the set of three modalities, it is possible to construct groups of two-modality sets in three different ways. For each dimension of a state modality, we can have a person-by-attribute set. For each person dimension, we can have an attribute-by-state set. For each attribute, we can have a person-by-state set. As will be seen later, we may have sets consisting of more than three modalities.

Differing Terminology

The notion of a state modality is probably better recognized in the literature as the "occasion" modality. Tucker (1963) has considered a sophisticated mathematical model of what he has designated a three-mode factor analysis model. In this model, he uses the term "mode" in the same sense as we use "modality." His three modes are persons by tests by occasions.

The term occasion may be regarded as a special case of what we have called state. This latter designation is more in line with the terminology suggested by Cattell (1966), who implies that the designation of "occasion" is too restrictive since it connotes a chronological ordering of the dimensions of the modality.

Cattell also prefers the term set to refer to what we call modality, after Tucker's mode. He points out that etymologically the term mode is not logically appropriate for the concept. This objection is rather persuasive, but to accept his suggestion to use the term set instead of modality would require the choice of another word for what we have called a group of two or more modalities. One might consider the term group for what we have called set; however, it seems best at this point to reserve the word group for any specified collection of things and attribute to set a hierarchical status above modality.

One might be tempted to argue "a rose by any name" and question why all the concern over terminology. But, as we shall see, even with the best of care in the selection and definition of terminology, the construction of a tightly structured model of expectancy theory can encounter some very slippery issues. Actually Cattell (1966) has proposed a most ambitious and fascinating system of data representation that he refers to as the data box. He appears to believe that it is adequate to accommodate any viable model of human behavior that one might formulate. It is still not obvious, however, how some of the emerging concepts of a comprehensive expectancy theory can be adequately accommodated by the Cattell data box. This is not to say, however, that the effort might not be eminently worth while. However, for the time being, we shall proceed over an as yet imperfectly charted sea and hope to work out a reasonably useful charting as we go along.

Before leaving the issue of an appropriate name for what we have called modality and what Cattell wishes to call set, we should note that the term "dimension" has been

suggested. The person-by-test-by-occasion matrix has sometimes been called a three-dimensional matrix. This appears to be an unfortunate pirating of the word dimension. The term has a long tradition in geometry, algebra, and statistics as referring to the group of things included in a modality. It is common in psychometrics to speak of the people or the test space where the space is characterized as n-dimensional. In fact, the term "space" is used in the same sense as we use "modality" and it might even be argued that it would be a preferable designation. However, because the term carries such a large burden of connotations that are irrelevant if not counter to our purposes, it is rejected.

Levels

We suggest then that any scientific model must begin with a data set consisting of at least two modalities. To be nontrivial, at least one of the modalities should have two or more dimensions. Each dimension of one or more of the modalities must have two or more levels.

The levels of a dimension may vary from two to infinity. The special case of a two-level dimension for which the levels are 0 and 1 is called a binary variable. If the number of levels is infinite, the dimension is called a continuous variable.

Confusion in the Use of Terms

Levels can play an important role in expectancy theory. The unfortunate practice of using the term in two different senses is common in expectancy theory literature. The concepts of "level of effort" and "level of performance" play a key role in most of the expectancy theory models. Although the words are used here in the same way we have defined them, they may also be used in connection with types of outcomes. Writers refer to "first level outcomes" and "second level outcomes." If the usage of "level" in connection with effort or outcome is accepted, then we must by all means proscribe its use to distinguish between types of outcomes. This is the course we adopt.

Levels as a Dimensional Group

There is a more significant issue with reference to the concept of levels of dimensions that we must develop. It is the notion that the levels of a dimension may themselves be regarded as a dimensional group. A dimension may be expanded into its levels to constitute a group of binary dimensions. Although this concept is not new in traditional multivariate analysis models, it is not commonly used. The eta-coefficient, which is used to test linearity of regression, can be shown to be the multiple correlation coefficient between a dependent variable and a group of independent variables that is obtained by expanding a single independent variable into a group of binary variables. These binary variables consist of a specified group of levels of the independent variable. Actually the principle of using levels as variables is implied in typical analysis of variance and covariance models when discrete variables are polychotomized to obtain experimental designs.

In expectancy theory, the expansion of effort and performance dimensions into binary level dimensions assumes a major feature of the principal models. The emphasis of expectancy models on binary level dimensions obtained from expanding a dimension into its levels is an important distinction between traditional multivariate models and contemporary expectancy theory models. This emphasis need not cast doubt on the validity or viability of the models. It may just as well turn out to be an advantage of these models over traditional models and may suggest profitable extensions and elaborations of these established models.

Outcomes

A curious complication of the concept of the levels of a dimension arises in connection with certain types of events that play key roles in expectancy theory. One of these is the group of events called outcomes. Depending on the writer, these may be first or second level (sic) outcomes. There is no general agreement on just how an outcome should be defined operationally. The definition that "an outcome is simply anything that an individual might want to attain" is not very helpful and is in fact not followed in most expectancy theory formulations. The schema of Campbell, Dunnette, Lawler, and Weich (1970) and of Campbell and Pritchard (1976) agree in their postulation of two stages of outcomes. The former group calls the first stage "rewards" and the second stage "needs," while the latter refers to the first stage as "outcomes" and the second as "basic needs." These appear to be slight semantic variations. The essential point is that, in both cases, an outcome is considered as something that either does or does not occur.

Nebeker and Moy (1976), among others, have recognized that this conceptualization causes difficulties when one attempts to apply the concept of valence to an outcome. An outcome may be positively or negatively valent but what about the valence of not attaining an outcome? Does it make sense to speak of the valence of a nonoutcome or a "nothing?"

The issue can be brought into sharper focus and then induced to disappear completely if one first considers a group of outcome dimensions, each with its associated levels. At once, the occurrence of an outcome dimension becomes a level of 1 and its nonoccurrence, a level of 0. The logical difficulty has arisen through a failure to recognize explicitly that the concept of valence does not apply to a dimension of an outcome modality but rather to a level of a dimension of a modality.

This modest discovery reminds us of the obvious—that outcomes rarely occur in an all-or-none fashion but rather in various degrees (levels). A person does not necessarily get exactly either a \$50 reward for a good performance or no reward at all. He may get a \$10 reward or a \$500 reward.

As we have already seen, the concepts of expectancy and valence are associated in expectancy models. As we shall see later, a valence specification of an event is typically associated with a person's expectancy of the occurrence of the event. If we insist that the event whose valence is specified must be the level of a dimension, then theoretically a person should be willing and able to express his level of expectancy that a given level of the event will occur. This conclusion has pushed investigators to object that there are definite limits to the number of different levels persons can be expected to discriminate reliably with reference to both expectancy and valence.

The objection is well made and the limits are soon reached. Nevertheless, the issue raised is important. It has been resolved in part by restricting the model to two levels of outcome, as has been done by Moy and Nebeker—zero and one. This is the smallest number of levels possible.

As a matter of fact, the level issue is explicit in most expectancy theory research in the treatment of effort and performance or goal accomplishment events. The use of phrases like "high effort" vs. "low effort" and "good performance" vs. "poor performance" are common in expectancy theory literature and experimentation. Here level specification for the two types of events is clearly implied. Again, Moy and Nebeker (1976) have recognized that the flexibility and power of the expectancy theory model could be increased by increasing the number of levels to be considered in each type of event.

Determination of the Number of Levels

The foregoing discussion of levels of dimensions raises the issue of how the number of levels for any particular dimension is established. We have seen that, for multivariate analysis models in general, the number can vary from two to infinity. A consideration of the procedures, operations, logic, and whatnot by which the levels of a dimension may be determined could lead us into a vast area of theory and practice in the numerous special fields of each of a large number of scientific and technical disciplines. Much as the need exists for a comprehensive treatise on this subject, the preparation of such a treatise would constitute an ambitious project in itself, and we shall attempt to treat only the necessary bare essentials of the subject in this document.

The determination of the levels of a dimension is of great importance in multivariate analysis models in general and in expectancy theory models in particular. We began our discussion of multivariate analysis models with the following sequence of concepts: sets, the modalities within a set, the dimensions within a modality, and the levels within a dimension. Although this appears to be a logical and natural order in which to introduce these key concepts, as a matter of fact, it is only by means of observational and operational explication of a series of levels that a dimension can be defined. Similarly, a modality can be defined ultimately only in terms of dimensions. And finally, a set can be defined only in terms of its modalities.

This sequence of definitional processes places a heavy burden on the observational and operational definitions of levels. At the same time, if these definitions are clearly and unambiguously formulated, we can provide a solid foundation for the definitions in the ascending hierarchical structure of basic concepts. This point cannot be too strongly emphasized. Psychologists and other social scientists agonize at length over the meaning and definition of words. Much of this achieves no more than masochistic semantic self-gratification. In the last analysis, no concept can mean any more or less than that given it by the observational and operational definitions of the levels on which it is based, either directly or indirectly. This is another way of saying that the mechanics of quantification define a variable or other concepts derived from it.

Measurement Scales

No discussion of levels of dimension is complete without at least a nod of recognition to the traditional classification of measurement scales, together with a brief critique of them.

1. Nominal scale. This is usually treated as the most primitive of scales and is simply meant to indicate the existence or presence of an event. Sometimes this scale is referred to as "categorical" or "qualitative." Most discussions of this classification are not notably sophisticated, even when conducted by sophisticated analysts. Usually the scale is not explicitly recognized as a special case of a two-level dimension, which indeed it is.

2. Ordinal scale. This scale posits a series of number, x_1, x_2, x_3 , etc., but makes no assumptions beyond the one that x_1 is less than x_2 , x_2 is less than x_3 , etc. An example of an ordinal scale would be: never, rarely, frequently, always. Presumably, "never" is less than "rarely," "rarely" is less than "frequently," etc. However, in general one cannot say that the difference between "never" and "rarely" is the same as the difference between "rarely" and "frequently"; the same is true for the other adjacent adverbs of degree. If

the members of a group of things are rank-ordered according to some specified attribute or dimension, the rank orders may be said to constitute only an ordinal scale. The essential characteristic of an ordinal scale can be expressed by $x_{i+1} - x_i \begin{matrix} < \\ > \end{matrix} x_i - x_{i-1}$, where the range of i is over the number of intervals. This means that the scale intervals are not all necessarily equal.

3. Interval scale. This scale differs essentially from an ordinal scale in that the intervals of the scale are all equal. The difference between 2 and 3 is the same as the difference between 3 and 4, etc. The essence of the scale can be indicated symbolically by $x_{i+1} - x_i = x_i - x_{i-1}$. This representation is the same as for the ordinal scale except that the "less than" (<) and "greater than" (>) symbols have been deleted from above and below the equality sign respectively.

4. Ratio scale. A ratio scale has an absolute zero point, unlike the interval scale which says nothing about the zero point of the scale. In an interval scale, it is not necessarily true, for example, that 40 is twice as much as 20. In the Fahrenheit temperature scale, 100 degrees cannot be said to be twice as warm as 50 degrees. Converted to Centigrade, these two temperatures would be approximately 38 degrees and 10 degrees respectively. Yet one would obviously not say that the higher temperature is about four times as great as the lower. A ratio scale, on the other hand, is one which has, according to some meaningful rationale, an absolute zero point.

5. Absolute Scale. Suppose we do have an interval scale with an absolute zero point for both height and weight so that we could say that a man is 70 inches high and weighs 140 pounds. Obviously, we would not say he is twice as heavy as he is tall. The problem clearly is that height and weight have not been expressed in absolute units of measurement. In general while we can sometimes define absolute origins of measurement, the definition of absolute units is usually arbitrary or, at best, is based on systems of rational considerations.

We shall have to come back to the problem of determining levels of dimensions when we undertake the formulation of a more comprehensive expectancy theory model. The problem is, however, a crucial one even in the more traditional models of multivariate analysis. The traditional system of scale types that we have discussed does not play as crucial a role in psychometric models, or even quantitative models in other scientific disciplines, as the discussion in the literature on measurement might lead one to believe. The numbers that are usually generated in experimental or simple observational or recording situations usually come to hand without serious questioning as to their scale-type classification.

To take one example, let us consider a test score on an objective arithmetic test. In the simplest case, it may be the number of items answered correctly. But actually we may regard each item in the test as a dimension of the test. We may consider each item dimension as having two levels--1 for a correct answer and 0 for an incorrect answer. The test score is then the sum of level 1 responses for the items in the test. Few would be concerned whether the score is from a nominal, ordinal, interval, ratio, or absolute scale. Some might suggest that the score on an individual item is from a nominal scale but, until further questions are raised, such a suggestion has doubtful value in attempts to classify the test score in terms of the traditional scale type. Perhaps the most significant thing about an item score is that it is a two-level or binary type and this type is of great importance in certain multivariate analysis models.

The foregoing is not meant to imply that test scores or other level numbers should always, or even usually, be taken at face value. A vast statistical literature is concerned

with the derivation of meaningful numbers from raw scores, such as standardized, normalized, or other types of transformed measures. The issue of how levels of dimension or scales in general should be generated cannot be dismissed lightly, but the goal of absolute scale-type values for all levels utilized in multivariate analysis models, either currently available or to be developed, is neither realistic nor necessary.

For many years, a great body of theory and technology has been accumulating in the field of scaling techniques. The goal is to utilize nominal or ordinal type observations to generate at least interval-type measures and hopefully even ratio, if not absolute, type measures. The so-called nonmetric models of multidimensional scaling due to Shepard (1962), Kruskal (1964), Coombs (1964), Guttman (1968), Lingoes (1965), and others are really attempts to generate rational metrics from nominal or binary-type observations. Some of the more forward looking and innovative expectancy model proponents may be tempted to incorporate some of these more recent efforts into their models. It is possible, however, that currently, and perhaps for the future as well, the multidimensional scaling models can be only a false hope that will divert their energies from more profitable directions for improving the power of the expectancy theory model.

DIMENSIONS OF A MODALITY

Levels

We have dwelt at length on the level concept and we shall be returning to it later when we take up the expectancy model in greater detail. Now we must consider further aspects of the dimensionality concept. We have already seen that a dimension of a modality can be expanded into a group of level dimensions, provided it has more than two levels. Each of these level dimensions has two and only two levels—0 and 1. This procedure for proliferating dimensions is not unknown in classical multivariate analysis models, has played a crucial role in expectancy theory models, and can be expected to play an even more important role in further elaborations of the latter models.

Manifest vs. Latent Dimensions

We shall now consider the dimensions concept in more detail. Most statistical models, or indeed the data matrices for which they are used to characterize the dimensions of the modalities in a two-modality set, are dictated or specified by the overall research project and the real-world setting of the project. For example, we may wish to construct a data matrix of test scores and other attribute dimensions on each of 20 different tests and perhaps other performance and physiological measures for each of 100 different persons. We shall refer to the 20 different measures as the manifest dimensions of attribute modality and to the 100 persons as the manifest dimensions of the person modality. Suppose, however, that we have a technique for finding another set of attribute dimensions, each with its level assignment for each person such that, by a predetermined arithmetical set of procedures, we can reproduce the original data matrix.

Suppose that this derived attribute modality of dimensions possesses properties that, in some specified sense, are markedly simpler than the original data matrix with the manifest dimensions of the attribute modality. Actually such arithmetical or, more generally, mathematical procedures do exist and are represented by the factor analytic models. We shall refer to this derived modality of attribute dimensions as latent dimensions of the attribute modality. In general, we may speak of the latent as well as the manifest dimensions of a modality. As will be seen, the concept of latent dimensions of a modality has important implications for expectancy theory model building.

Predictor vs. Criterion Dimensions

Another dimensional concept that is fundamental to most traditional multivariate analysis models and to expectancy theory models is that of criterion vs. predictor dimensions within the attribute modality. Most mathematical statisticians refer to these two types of dimensions as dependent and independent variables respectively. This latter designation is perhaps unfortunate, not only because of the connotations it carries for random vs. fixed variables but also because it fails to emphasize the fundamental objectives of pure and applied science; namely, the prediction and control respectively of events. The designations favored by mathematical statisticians are of doubtful utility for models that strive for the prediction and control of human behavior.

Variables

We have used the word variable without providing an explicit definition of it. It is used so freely in scientific and quasiscientific discussions that its meaning is usually taken for granted. However, before concluding our preliminary discussion of predictor and criterion variables and continuing our discussion of types of dimensions in general, we shall define the word in terms of the concepts we have already introduced.

We shall define a variable as a dimension together with its specified levels. Thus, if a dimension has only two levels, it is called a binary variable. We shall take great pains not to refer to such a variable as "qualitative," "categorical," "nominal," "dummy," or any other of the terms that are used to refer to binary type variables. If the levels of a dimension are specified to be greater than 2 but finite in number, we shall refer to the dimension with its specified levels as a discrete variable. If the number of levels of a dimension are specified to be infinite and their limits are finite, we shall refer to the dimension with its specified levels as a continuous variable.

In general, we may speak of n -ary variables, where n refers to the number of levels of a dimension. Thus, n is 2 for a binary variable. For a discrete variable, n is greater than 2 and less than infinity. For a continuous variable, n is infinite and its limits are finite.

Scales

We note in these definitions that none of the types of scales that we have discussed has been invoked with the exception of nominal scales that we treat as synonymous with binary variables. Although it might be interesting to attempt to integrate the concepts of scale types and generalized n -ary variables, no serious efforts appear to have been made thus far. The multidimensional scaling models are more closely associated with scale types than with the concept of generalized n -ary variables. However, for a comprehensive expectancy theory model, the concept of n -ary variables is believed to have more relevance than that of scale types.

Predictor and Criterion Variables

Having hopefully clarified the concept of variables, we shall now conclude briefly our preliminary discussion of predictor and criterion variables. In general, criterion variables are defined as those we wish to predict and predictor variables, as the ones on which the predictions are based. The literature that attempts to clarify the distinction between predictor and criterion variables is extensive (see Horst, 1966) and will not be reviewed here. But the concepts are fundamental to expectancy theory models and unfortunately they have not been clearly recognized in some of the literature.

Two major distinctions characterize the predictor-criterion dichotomy. First, the criterion variables are typically regarded as having significant social value in their own right while predictor variables need not necessarily have value in themselves. For example, a career success variable would be regarded as being socially significant, whereas performance on an aptitude test would not of itself be socially significant. Only its relevance for predicting career success would be of primary interest.

A second distinction between predictor and criterion variables is chronological sequence. Typically, predictor measures are available before criterion measures. A test score is available for a person before he actually achieves a degree of success in a job or career activity. These and other aspects of the predictor-criterion distinction will be further considered when we take up the expectancy theory model in greater detail.

The Interaction Dimension

We have considered three different kinds of dimensions of the attribute modality: (1) dimensions that are themselves levels of the given attributes, (2) manifest vs. latent dimensions, and (3) predictor vs. criterion dimensions. A fourth type of dimension that is important for multivariate analysis models in general and for expectancy theory models in particular will be called an interaction dimension.

Definition

An interaction variable is the product of two or more other variables. An interaction dimension is a dimension whose levels are products of levels of two or more other dimensions. The interaction variable plays an important role in analysis of variance and covariance models and in most traditional experimental design formulations.

Moderator Variables

The interaction variable has also received prominence in multiple regression models in recent years under the guise of moderator variables. Moderator variables are simply second-order interaction variables or variables generated from the products of two other variables. It is perhaps unfortunate that this term has crept into the literature and has assumed such a conspicuous role. The emphasis on the terminology has unwarrantedly given status to second-order interactions at the expense of possibly equal or more important higher-order interactions.

Generalized Models

Generalized models of interaction binary variables have been developed under the name of configural analysis. The work of Lazarusfeld (1950) and others on latent class analysis is closely related to interaction binary variable models. A more comprehensive treatment of interaction variable models is given by Horst (1973).

Important Role of Interaction

Interaction plays an especially important role in current expectancy theory models and may well continue to do so in a more comprehensive model of the theory. The concept of interaction variables is particularly significant in expectancy theory models because of the questions concerning their validity that are constantly being raised in the literature. Some of these questions arise in connection with the concept of level-type dimensions, which, as we have seen, can be generated from single multilevel dimensions.

Concerns Over the Use of Interaction Variables

Some of the concerns expressed over the use of interaction variables in expectancy models are not peculiar to these models but apply to other traditional multivariate analysis models as well. We shall also see that other concerns that are raised about various aspects of the expectancy theory models are the same types of concerns that have been raised about multivariate analysis models in particular and scientific models in general. A failure to recognize this fact and a lack of understanding of how investigators have come to terms with, or effected resolution of, these concerns has generated undue anxiety in the expectancy theory literature. We shall examine some of these issues in some detail when we review the slated postulates of expectancy theory and other methodological issues that have been enumerated.

Linear vs. Nonlinear Relationships

The issue of types of relationships among variables will be considered in greater detail later. We note that the conventional distinction is between linear and nonlinear relationships. This subject cannot be intelligently discussed without considering what it is that is being related. Although the relationship is typically between two or more variables, there is frequently confusion about what it is that is linearly or nonlinearly related. Sometimes an interaction variable is regarded as an example of a nonlinear relationship. But one may have linear relationships among interaction variables. It is only when the concept of parameters in addition to variables is introduced into a model that the distinction between linear and nonlinear relationships becomes methodologically and analytically crucial. The parameters of a model may involve nonlinear functions.

Expectancy theory literature has not clearly distinguished between additive vs. multiplicative combinations of levels on the one hand and linear vs. nonlinear relationships on the other. Neither is this distinction always clear in other multivariate models. We shall discuss this distinction in more detail when we take up the subject of parameters in the multivariate model.

Intervention Variables

A fifth type of dimension from an attribute modality has assumed ever greater importance in the practical applications of multivariate analysis models and must also play a central role in viable expectancy theory models. The variable that includes this type of dimension has been called intervention variable, manipulable variable, control variable, and experimental variable. This type of variable is of great practical importance because it is on the basis of such variables that applied scientific disciplines, including applied psychology, purport to exercise control or influence over events, whether these events be the nuclei of atoms, people, nations, or ideological blocks of nations.

The Importance of Intervention Variables

The importance of these manipulable variables has been highlighted in recent years by the billions of dollars that the U. S. Congress has appropriated in the hope of improving methods and procedures in our educational institutions, particularly as these relate to the so-called disadvantaged segments of our population. This type of variable is also crucial in the application of expectancy theory models to organizational behavior, whether in industrial, educational, governmental, or military settings. It is central to any social system where the mission of the system is clearly defined and where efficient execution of the mission is regarded as crucial. In fact, it may be said that intervention-type

variables are important for the operation of any social system or subsystems thereof, irrespective of the ideological orientation of the system.

The Special Case of Predictor Variables

In general, intervention variables may be regarded as a special case of predictor variables. If variables can be found that can be manipulated to influence criterion variables such as socially desirable behavior, it may be possible to influence such behavior through the manipulation of these variables. In an organizational setting, manipulable variables may be sought by means of which individual performance and, through this, group performance may be influenced so as to increase the attainment of the mission objectives of the organization.

A special case of an intervention variable is found when we establish an experimental or treatment group of persons or other entities to test whether a special treatment of the experimental group is followed by a significantly better prespecified outcome in the experimental group than in the control group. In this example, the treatment variable is the intervention variable. It is a binary variable in which the members of the experimental group are assigned a level of 1 and the control group, a level of zero. Such a variable is a binary, predictor, manipulable variable, according to the terminology we have suggested. In addition to this variable, the simplest form or the model also implies another pretreatment n-ary variable and a posttreatment n-ary variable.

N-ary Predictor Variables

The pretreatment variable is also a predictor variable. The posttreatment variable is a criterion variable that quantifies the socially desirable results that it is hoped to achieve. A great deal, both sophisticated and naive, has been written about specification of the n-ary predictor variable (sometimes called "variate") and its role in the assignment of persons or other entities from the entity modality to the so-called treatment and control groups.

Practical Suggestions for Intervention Variables

The literature on intervention variables is too extensive, the issues too complex, and the proposals for their resolution too controversial to discuss even superficially in this document. However, we must emphasize at this point that (1) intervention variables in general need not be restricted to binary variables, (2) the levels of these variables can be assigned to dimensions of an entity modality by agencies independent of the entities themselves, and (3) the rules and procedures for these assignments can validly be much more flexible than those typically imposed by hard-core experimentalists. These three points are briefly illustrated below:

1. The intervention variable may be individual instruction. The amount of individual instruction may be varied according to the achievement test scores of a group of fourth-grade pupils. It need not be the same for all achievers below a given score and absent for all those above.

2. The decision that individual instruction be made available to the student is not made by the student himself but, rather, by a teacher or some other person.

3. In research to test whether individual instruction increases achievement levels, it is typically insisted that a total group of students first be divided into two groups exactly equal in achievement and that one of the groups (experimental group) be given the

treatment (individual instruction) and the other group (control group) be denied the treatment. The achievements of the two groups are compared after the treatment program and tested to see whether there is a significant difference.

Several observations are in order with reference to this third point. First, it should be noted that the requirement of comparability of the two groups before treatment is not a strict requirement for testing the effectiveness of the treatment. Statistical techniques are available for investigating the effectiveness of the treatment, even though the two groups are not comparable with respect to the proficiency that is under consideration.

A second observation is that we must distinguish between projects designed to apply intervention procedures and those designed to test their effectiveness. This distinction is important because the U. S. government has expended enormous resources over the years in the application of intervention procedures without first conducting research to determine the effectiveness of the procedures. This untested intervention procedure has gone by the name of "demonstration projects," while the subsequent procedures to test their effectiveness are called "evaluation projects." We shall not attempt to document this charge as the evidence for it abounds on all sides. If one examines budget allocations in the area of education alone, the ratio of funds allotted for evaluation to those allotted for demonstration is roughly 1:50. In other words, we spend vast sums on projects whose social value is unknown and then we spend paltry sums to test their value. What we should be doing is to spend modest or even generous sums to test the utility of social programs before huge sums are committed to them.

Experimentation-utilization

This apparent digression has been made to emphasize that the intervention type variable is not only of considerable theoretical value, but its great practical significance for the control of human events is sufficient justification for according it a special classification as a type of predictor variable. It should be clear by now, however, that its role as a predictor variable can be established only by experimentation and its use as a predictor variable is justified only after experimentation has established its value as a predictor of a criterion variable. This sequence of experimentation and utilization is of course applicable to predictor type variables in general.

Endogenous-exogenous Predictor Dimensions

We shall consider briefly one more type of attribute dimension distinction that is relevant to multivariate analysis models in general but perhaps more particularly to expectancy theory models. This distinction is between endogenous and exogenous predictor dimensions. The distinction is made between those dimensions that characterize directly and internally the structure and functioning of the entity (person) and those variables that are external to the organism but associated with it.

Endogenous attribute dimensions of a person may include a wide variety of physiological characteristics. They may also include different aptitude dimensions. Acquired varieties of skills and proficiencies would be included in endogenous dimensions, as would personality, emotional, attitudinal, and interest dimensions.

Exogenous attribute dimensions may involve all sorts of background and personal history variables. Biographical and socio-economic data, both past and present, would be included. In general, one might say that exogenous variables include all those environmental factors, past and present, that might be regarded as having some impact on the organism.

We might say that endogenous variables are those that characterize past or present states of the organism. These might also be called organismic variables when the dimensions of the entity modality consist of biological organisms. The exogeneous variables might also be called environmental variables. These designations of "organismic" and "environmental" become especially appropriate when we attempt a formulation of the expectancy theory model that gives major emphasis to the concepts of response patterns of the organism and stimulus patterns of the environment. This distinction should help to clarify some of the ambiguity and confusion found in current expectancy theory formulations.

RELATIONSHIPS AMONG VARIABLES

We have considered at some length the basic structural elements of multivariate analysis models in general. We have also, incidentally, included occasional references to relationships among some of these elements. Specifically, we have referred to the relationships between variables. A variable was defined as a dimension with its pre-specified levels. The concept of variables and their relationships with one another is fundamental to all scientific models, including multivariate analysis models and expectancy theory models. This concept is central to the prediction and control of all natural phenomena. It is particularly important in expectancy theory models that we understand clearly the basic concepts involved in the functional relationships among variables.

In our discussion of the kinds and types of variables, we considered primarily the attribute modality dimensions and their associated levels. We discussed briefly occasion, condition, and evaluator modalities. We referred more frequently to the entity or person modality. Thus, we have placed greater emphasis on the entity-attribute set of modalities than on other sets of modalities. The reason for this emphasis is that these two modalities must play a key role in any definitive discussion of the functional relationships among variables. To better understand their role, we shall consider first the entity modality.

The entity modality is fundamental to all other modalities. It is the modality of primary interest to which other modalities refer. Most obviously and directly, the dimensions of the attribute modality refer to the dimensions of the entity modality, whether these latter be persons, geographical regions, samples of physical or chemical substances, or whatnot.

Without attempting to pursue all the ramifications of this statement, we suggest at this point that other modalities may be treated conceptually as attribute modalities. This concept has been developed at greater length elsewhere (Horst, 1963, 1965). We shall see now how we may articulate the entity-attribute modality concept with the more traditional mathematical concepts of functional relationships among variables.

The Simple Variable Function

First we start with the very simple equation:

$$y = f(x).$$

This equation says that y is a function of x . Special cases of (11) may be

$$y = x^2 \tag{12a}$$

$$y = x + x^2 \tag{12b}$$

$$y = \log x \quad (12c)$$

$$y = 1/x \quad (12d)$$

We have said nothing about the value of x in equations (11) and (12). Suppose that we regard x as a dimension of an attribute modality and we have an entity modality consisting of 10 person dimensions. Let us assume that the attribute dimension x has four levels (1, 2, 3, 4) and that one of these levels is assigned to each of the 10 persons as follows:

| | | | | | | | | | | |
|------------|---|---|---|---|---|---|---|---|---|----|
| Person No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| x - Level | 1 | 3 | 2 | 4 | 1 | 2 | 2 | 4 | 3 | 1 |

Figure 1.

We now have an example of a 4-level dimension of an attribute modality, and a 10-dimension entity modality to each of which is assigned a level of the attribute dimension x . Suppose next we have another dimension of an attribute modality that we shall call y and that this dimension has 16 levels, thus: (1, 2, ..., 16). From these levels, one is assigned to each of the 10 person dimensions. Assume that the levels are assigned as follows:

| | | | | | | | | | | |
|------------|---|---|---|----|---|---|---|----|---|----|
| Person No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| y - Level | 1 | 9 | 4 | 16 | 1 | 4 | 4 | 16 | 9 | 1 |

Figure 2.

We may now consolidate the two tables in columnar form, thus:

| Entity Modality | Attribute Modality | |
|-----------------|--------------------|---------|
| Person No. | y-level | x-level |
| 1 | 1 | 1 |
| 2 | 9 | 3 |
| 3 | 4 | 2 |
| 4 | 16 | 4 |
| 5 | 1 | 1 |
| 6 | 4 | 2 |
| 7 | 4 | 2 |
| 8 | 16 | 4 |
| 9 | 9 | 3 |
| 10 | 1 | 1 |

Figure 3.

Figure 3 is a simple data matrix whose column dimensions y and x constitute the attribute modality and whose row dimensions 1 to 10 constitute the entity modality. We let equation (12a) be a particular function of equation (11) so that a prespecified value of x is squared to give the corresponding value of y . We see then that, in Figure 3, an x level is squared to get a y level.

If we started with the data matrix in Figure 3, we could by inspection deduce that the y variable was obtained by squaring the corresponding x value and so could have expressed the general functional relationship given by equation (12a). But, having deduced this equation from the data matrix, we might reasonably infer several things. First, suppose that, for additional entity dimensions, we had for each a level assignment from one of the four x levels but no level assignment for the corresponding y dimension. We might reasonably assume that the appropriate y level could be obtained by squaring the given x level. We could also assume that, if the number of x levels were increased, equation (12a) would still be used to determine unspecified y values from given x values. In fact, if the number of levels were increased without limit to generate a continuous variable x , we could use equation (12a) to determine a y value corresponding to any one of the infinite possible number of x values. We would then have arrived at the general mathematical equation giving y as a function of x in (12a).

This discussion attempts to show how a simple mathematical equation can express the relationship between two attribute dimensions in the data matrix of ten entities, if indeed such a relationship exists. It is obviously a great oversimplification of data obtained from real-life situations. We shall therefore introduce another concept into equation (12) by writing

$$y = a x \quad (13)$$

where the new concept is represented by the letter a .

Suppose again we have a number of different entities for which x values are available and also corresponding y values. Suppose we observed that, for every x value, the corresponding y value was twice as large. We could then decide that, based on available data, the value of a in equation (13) is 2 and we can write

$$y = 2 x. \quad (14)$$

In passing, we note that in equation (14) y is a linear function of x , while in equation (12a) y is a nonlinear or curvilinear function of x . This means that, if we plot y values against x values for (14), we get a straight line, while if we do the same for equation (12a) we get a curved line. Another important difference in the two equations is that equation (13) has both a variable x and an unspecified constant a on the right side of the equation while equation (12a) has only the variable x in it. The value x may vary from one entity to another while the constant a remains the same for all entities.

The Two-parameter Model

We may consider still another function y of x in equation (12) that is slightly more complicated than equation (13); thus,

$$y = a x + b. \quad (15)$$

On the right side of (15), we have the variable x and two unspecified constants a and b . The first of these is a value by which the variable x is multiplied and the second, b , a value that is added to the product. Again, the value of level of x may vary from one entity dimension to another but both the values a and b remain constant for all values of x . As in equation (13), the variable y is a linear function of x .

Suppose now we go back to page 10 and recall the distinction between the estimate of a value and the value itself. Let us call \tilde{y} the estimate of y ; instead of (15), we write

$$\tilde{y} = a x + b. \quad (16)$$

Equation (16) is just like (15) except we have replaced y with \tilde{y} . Then we can write

$$\epsilon = y - \tilde{y} \quad (17)$$

where ϵ is the difference between the actual value of a y value and \tilde{y} is the value estimate for it from its corresponding x value by means of equation (16). There will be an ϵ , a y , and an x value for each entity in the entity-attribute set but, since we have not specified the constant values a and b , we do not actually know the \tilde{y} values nor the ϵ values.

Now equation (15) may be called a mathematical model of a multivariate system. We see that the model involves variables that include dimensions (with their levels) of the attribute modality and implies the dimensions of the entity modality. It also includes unspecified constants. Finally, it includes an explicit form of the relationship or function that y is of x and the unspecified constants a and b .

The Loss Function

Next we note that from (16) and (17) we may write

$$\epsilon = y - (a x + b). \quad (18)$$

For the i th entity, we may write

$$\epsilon_i = y_i - (a x_i + b). \quad (19)$$

If there are n_i entity dimensions, we may define another function:

$$\Phi = \Phi (\epsilon_1, \epsilon_2, \dots, \epsilon_{n_i}). \quad (20)$$

Equation (20) means that Φ is some function of the n_i ϵ values as yet undetermined.

Now with experimental data we do not expect the estimated and actual values of y (the criterion) to be equal for all n_i entities but we would like the differences to be as small as possible. The traditional least-squares approach is to determine the as yet unspecified a and b values so that the sum of the squares of the ϵ s, or residuals, shall be a minimum. The mathematical and computational procedures for accomplishing this are of course well known and perfectly straightforward. We write

$$\Phi = \sum_{i=1}^{n_i} \epsilon_i^2. \quad (20b)$$

The function Φ is called the loss function of the model.

Now, because of (19) and (20), Φ is a function of both the specified x_i and y_i values and the unspecified a and b values. Whatever the loss function, the values a and b are treated as variables that are determined so as to optimize the loss function. They are called the parameters of the model. We will consider these further shortly in a more general treatment of a multivariate analysis system because they can have an important role to play and have been almost totally ignored in the literature on expectancy theory models.

The Constraints

First, we must introduce another multivariate systems concept that can also have important implications for an adequate expectancy theory model. Suppose that, in addition to optimizing the loss function, we may also wish to determine the a and b parameters so that some other condition is satisfied such that the sum of the cubes of \tilde{y}_i shall be some prespecified value. We indicate the constraint by

$$F = F(a, b). \quad (21)$$

In the example suggested

$$F = \sum_{i=1}^{n_i} \tilde{y}_i^3. \quad (22)$$

Then we will write another function

$$\Psi = \Phi + \lambda F \quad (23)$$

where λ is a constant known as a Lagrangian multiplier. These Lagrangian multipliers can play a most important role in multivariate analysis models whenever it is desired to introduce one or more constraints on the parameters of the model in addition to the optimized loss function.

Now both the functions Φ and F in (23) above involve all of the x_i and y_i values as well as the parameters a and b . We could then rewrite (20) as

$$\Phi = \Phi(x_1, \dots, x_{n_i}, y_1, \dots, y_{n_i}, a, b) \quad (24)$$

and (21) as

$$F = F(x_1, \dots, x_{n_i}, y_1, \dots, y_{n_i}, a, b). \quad (25)$$

In (24), the exact mathematical function on the left depends on the loss function of the residuals that it is desired to optimize. Similarly, in (25), the exact mathematical function depends on the particular constraint that it is desired to satisfy.

Solving for the Parameters

In any case, the formulation of both the loss function and the constraining function is solely for the purpose of determining the values of the unspecified constants or parameters a and b .

The solution for these parameters proceeds as follows: We take the partial derivatives of Ψ with respect to a and b , respectively, thus:

$$\frac{\partial \Psi}{\partial a} = \frac{\partial \Phi}{\partial a} + \lambda \frac{\partial F}{\partial a} \quad (26)$$

$$\frac{\partial \Psi}{\partial b} = \frac{\partial \Phi}{\partial b} + \lambda \frac{\partial F}{\partial b} . \quad (27)$$

We let

$$\frac{\partial \Phi}{\partial a} = \Phi_a' \quad (28)$$

$$\frac{\partial F}{\partial a} = F_a' \quad (29)$$

$$\frac{\partial \Phi}{\partial b} = \Phi_b' \quad (30)$$

$$\frac{\partial F}{\partial b} = F_b' . \quad (31)$$

We equate the partial derivatives of the left sides of (26) and (27) to zero, thus:

$$\frac{\partial \Psi}{\partial a} = 0 \quad (32)$$

$$\frac{\partial \Psi}{\partial b} = 0 . \quad (33)$$

From equations (26) through (33), we get

$$\Phi_a' + \lambda F_a' = 0 \quad (34)$$

$$\Phi_b' + \lambda F_b' = 0 . \quad (35)$$

We recall now that λ is an unknown Lagrangian multiplier. However, we can eliminate it between equations (34) and (35) and get

$$\Phi_a' F_b' - \Phi_b' F_a' = 0. \quad (36)$$

We need not go into the mathematical details of how equation (36) is obtained. We simply observe that the left-hand side is a function of the unknown parameters a and b whose form may be determined from originally specified Φ and F functions. We may therefore

write (36) more simply symbolically as a function G of the unknown parameters a and b , thus:

$$G(a, b) = 0. \quad (37)$$

The constraining equation (21) contains only the two unknown parameters. Therefore, we have the two equations (21) and (37) from which we solve for the two unknown parameters a and b .

Summary for the Two-parameter Case

Let us now recapitulate what we have done.

1. We begin with a mathematical model consisting of variables and parameters.
2. We have n_i special cases of the model for each of the n_i dimensions of the entity modality. These are often called the observational equations.
3. We construct a loss function from the known values of the x and y variables for the n_i persons and the unknown parameters.
4. We impose some specified constraining function of the parameters.
5. We set up a total function that is the sum of the loss function and some unknown constant times the constraining function.
6. We take the partial derivatives of the total function with respect to each of the two parameters and equate each to zero, thus getting two equations in three unknowns, the two parameters, and the Lagrangian multiplier.
7. We eliminate the Lagrangian multiplier between the two equations, thus getting a single equation in the two unknown parameters.
8. We use the single equation from point 7 above the constraining equation from point 4 above to solve for the two unknown parameters.

We have still not completed the overall analysis procedure appropriate to most, if not all, multivariate analysis models. Several additional steps may be taken.

The parameters, now known, can be substituted back in the loss function Φ to calculate its actual value. By making certain more or less plausible assumptions and using more or less appropriate statistical tests, we can then arrive at a more or less valid decision as to the appropriateness of our model. But since these further steps involve somewhat more controversial procedures than do the first 8 we have listed, we shall not discuss them further at this time.

The Canonical Correlation Model

The specific example of a two-parameter model we have just discussed is a highly simplified special case of a multivariate analysis model. It is obviously a special case of the multiple regression model. In this example, we have one independent (predictor) and one dependent (criterion) variable. In the more general case, we have one criterion

variable and m predictor variables. In the special case we have 2 parameters, one of which (b) is the additive parameter and the other (a), the multiplicative parameter. In the general case, we have $m + 1$ parameters, one of which is the additive parameter and m of which are the multiplicative parameters, one for each of the predictor variables. In the traditional multiple regression model, it is not customary to introduce constraints to the solution for the parameters as we have done for the case of a single predictor and a single criterion variable. It was introduced in the simple case to illustrate its role in other more complicated models.

One of these more complicated models is what has come to be called the canonical correlation model. This model begins with a set of two modalities, an entity and an attribute modality. The entity modality has n dimensions. The attribute modality has m dimensions. Each of the m dimensions has two or more levels. The m dimensions of the attribute modality are divided into two groups, one of which has m_x dimensions and the other m_y dimensions. We call the dimensions with their levels in the m_x group the x variables and those in the dimensions with their levels in the m_y group the y variables.

The Parameters

For every entity i , we have a group of m_x values (x_{i1}, \dots, x_{im_x}) and a group of m_y values (y_{i1}, \dots, y_{im_y}). We also have a single group of m_x parameters (a_1, \dots, a_{m_x}) and a single group of m_y parameters (b_1, \dots, b_{m_y}).

We define a composite value u_i for each entity i , which is a weighted sum of the x_i values for the entity. The weights are the corresponding a parameters. For the general u value, we have then

$$u = \sum_{j=1}^{m_x} a_j x_j . \quad (38)$$

Similarly, we define a composite v_i value for each i entity, which is a weighted sum of the y_i values for the entity. The weights are the corresponding b parameters. For the general v value, we then have

$$v = \sum_{j=1}^{m_y} b_j x_j . \quad (39)$$

The Loss Function

Our mathematical model assumes that the a and the b parameters can be determined so that u and v are equal, or

$$u - v = 0 . \quad (40)$$

This model was first formulated by Hotelling (1936) and was originally referred to as "the most predictable criterion." Later it came to be called "the canonical correlation" model

but, like much of the terminology that evolves in scientific disciplines, this designation is not particularly appropriate.

The observation equations, or those corresponding to the entity dimensions, are of the form

$$\epsilon_i = \sum_{j=1}^{m_x} a_j x_{ij} - \sum_{j=1}^{m_y} b_j y_{ij}. \quad (41)$$

The loss function is

$$\Phi = \Phi (\epsilon_1, \dots, \epsilon_n). \quad (42)$$

In the traditional solution, the function Φ is taken to be the sum of squares of the ϵ 's so that

$$\Phi = \sum_{i=1}^n \epsilon_i^2 \quad (43)$$

The Constraints

We now specify two separate constraining functions, one for the a parameters and one for the b parameters. We specify that the a parameters shall be determined so that the averages of the squared u values in (38) are unity. Similarly, we specify that the b parameters shall be determined so that the average of the squared v values are also unity. These two constraints are given, respectively, by

$$\sum_{i=1}^n u_i^2 = 1 \quad (44)$$

$$\sum_{i=1}^n v_i^2 = 1. \quad (45)$$

But from (41) and (43), Φ is a function of the known x_{ij} and y_{ij} values and of the unknown a_j and b_j values. We can therefore write (42) in the more general symbolic form

$$\Phi = \Phi (a_1, \dots, a_{m_x}, b_1, \dots, b_{m_y}). \quad (46)$$

By the same logic we may also, because of equations (38) and (44), write the constraining equation

$$F_a = F_a (a_1, \dots, a_{m_x}). \quad (47)$$

Similarly, because of equations (39) and (45), we write the constraining equation

$$F_b = F_b (b_1, \dots, b_{m_y}). \quad (48)$$

Since in this example we have two constraining equations rather than a single one, we require two unknown Lagrangian multipliers λ_a and λ_b to set up the total function Ψ . We therefore have

$$\Psi = \Phi + \lambda_a F_a + \lambda_b F_b. \quad (49)$$

The Solution

To determine the a and b parameters that will optimize the loss function Φ and at the same time satisfy the constraining functions F_a and F_b , we first equate to zero each partial derivative of the function Ψ with respect to both the a and b parameters. This gives us a total of $m_x + m_y$, or m, new functions of the a's, the b's, and the two unknown λ 's. These we can write as

$$0 = \frac{\partial \Psi}{\partial a_j} = \Psi_{a_j}'(a_1, \dots, a_{m_x}, b_1, \dots, b_{m_y}, \lambda_a, \lambda_b) \quad j=1, m_x \quad (50)$$

and

$$0 = \frac{\partial \Psi}{\partial b_j} = \Psi_{b_j}'(a_1, \dots, a_{m_x}, b_1, \dots, b_{m_y}, \lambda_a, \lambda_b) \quad j=1, m_y. \quad (51)$$

With the two equations (47) and (48) and the m equations of (50) and (51), we can solve for the m unknown parameters and the two unknown Lagrangian multipliers. Again, we can also substitute the solved for a and b parameters into equation (42) to solve for the loss function Φ .

For this model, the derivative functions given by (50) and (51) are well known (Hotelling, 1936) as are the solutions for the a and the b parameters and for the two Lagrangian multipliers, λ_a and λ_b . As it turns out, there is in general more than one set of solutions but the one that gives an absolute optimum loss function is readily determined.

Significance Tests

Again, as in the two-parameter example, it is possible by making certain assumptions to get estimates of the appropriateness of the model.

An important distinction must be made, however, between how well the model fits the data corresponding to the entities on which the determination of the parameters was based and how well the model, using these same parameters, can be expected to fit data from another sample of entities (persons). A more difficult question to answer is how we can determine the parameters of a model from one set of data so that this set of parameters may be expected on the average to yield a minimum loss function on other sets. This question has received much less attention than the previous one although it is much more important.

The General Multivariate Case

We have considered a very simple two-parameter model and a much more complicated model known as the canonical correlation model. This latter model is of interest because it includes the multiple discriminant function model as a special case, while it in turn is a special case of what has been called the "m-set" model. We shall next consider the general case of a multivariate system consisting of (1) its structural elements and (2) the functional relationships involving structural elements. Suppose we have a set of

modalities, one of which is the entity modality and another an attribute modality. The attribute modality may be composed of groups of modalities, each with its dimensions and specified levels for each dimension. We assume level values available for each entity on each attribute dimension. We let n be the number of attribute dimensions. We let the attribute dimensions, together with their levels, be the variables (x_1, \dots, x_m) .

The Mathematical Model

Next, we consider a mathematical model consisting of the m variables together with s parameters (a_1, \dots, a_s) , which we express as

$$0 = f(x_1, \dots, x_m, a_1, \dots, a_s). \quad (52)$$

The i 'th observational equation corresponding to (52) would be

$$\epsilon_i = f(x_{i1}, \dots, x_{im}, a_1, \dots, a_s). \quad (53)$$

The Loss Function

The loss function Φ would be

$$\Phi = \Phi(\epsilon_1, \dots, \epsilon_m) \quad (54)$$

where Φ could be a least square, maximum likelihood, or other function to be optimized. From equations (52), (53), and (54), we could write more simply

$$\Phi = \Phi(a_1, \dots, a_s) \quad (55)$$

where it is understood that the constants x_{ij} are also involved in Φ .

The Constraints

Let us assume also that there are t constraints involving the s parameters and perhaps also the x_{ij} . These we indicate by

$$F_k = F_k(a_1, \dots, a_s) \quad k = 1, t. \quad (56)$$

We then set up the total function

$$\Psi = \Phi + \lambda_1 F_1 + \dots + \lambda_t F_t \quad (57)$$

where the λ 's are Lagrangian multipliers.

The Solution

Next we differentiate (57) partially with respect to each of the a 's and get equations of the form

$$\frac{\partial \Psi}{\partial a_j} = \frac{\partial \Phi}{\partial a_j} + \lambda_1 \frac{\partial F_1}{\partial a_j} + \dots + \lambda_t \frac{\partial F_t}{\partial a_j} \quad j = 1, s. \quad (58)$$

We assume all the partial derivative functions indicated in (53) can be found. Equating each equation of the type (58) to zero, we could write

$$0 = \Psi_j' (a_1, \dots, a_s, \lambda_1, \dots, \lambda_t) \quad j=1, s. \quad (59)$$

We have therefore s equations of the type (59) and t of the type (56) by means of which to solve for the s unknown parameters and the t unknown Lagrangian multipliers. Again, with certain assumptions, we can substitute the values of the parameters we have found into the loss function (55) as a basis for estimating the appropriateness of the model given by equation (52).

Significance Tests

We must now return to a question that was implicit in the conclusion of our discussion of both the special cases and the general case of a multivariate analysis system. This question concerns the confidence we can have in the appropriateness of the model we have chosen for our data.

First, we must consider three crucial values: the number of entities, the number of parameters, and the number of constraints. In the traditional multiple regression model we let N be the number of entities and m the number of independent variables. Then m is also the number of parameters.

Now if the number of observations or entities is equal to the number of parameters and there are no constraints, it is well known that the loss function is zero and the multiple correlation in the sample is equal to unity, irrespective of its value in the universe. (For the present, we shall use the terms "sample" and "universe" in the traditional usage without attempting either to defend or attack the concepts.)

If the number of observations is less than the number of parameters and there are no constraints, we can have an infinite number of sets of solutions for the parameters. Each of these will give a loss function of zero and a multiple correlation of unity. It is possible to require some function of the parameters alone to be optimized and regard the observation equations as constraints on the optimizing solution. In this way, a unique solution for the parameters can be found. Promising as this approach may be, it could lead us far afield from considerations more immediately relevant for expectancy theory models. If the number of entities plus the number of constraints is greater than the number of parameters, then in general the loss function will not vanish.

In general, significance tests may be defined in terms of the loss function, the number of observations, the number of constraints, and the number of parameters. If we use the term "conditions" to cover both observations and constraints and let N be the total number of the two, we may write

$$P = P(\Phi / \sigma, N, s) \quad (60)$$

where P is an estimate of the goodness of fit of the model in a new sample, Φ / σ is the ratio of residual to total variance, and N and s are as previously defined. As N approaches s , P approaches 0. As Φ approaches zero, P approaches unity. An example of a function that satisfies these conditions is

$$P = (1 - \frac{\Phi}{\sigma}) (\frac{N-s}{N+s})^2. \quad (61)$$

As either s or N approaches zero, the second parenthesis approaches 1. As N approaches s , this parenthesis approaches 0. Presumably σ must be equal to or greater than Φ .

Therefore, the right parenthesis must lie between one and zero. Also, P can vary only between zero and plus one. In general, mathematical statisticians give much more sophisticated treatments of this topic that are beyond the scope of this discussion.

Special Cases of Significance Tests

Suppose then we take the most general case for which we have the three special cases:

1. The total number of observations is much greater than the number of parameters. In this case, we may or may not impose additional constraints.
2. The total number of observations is less than the number of parameters. In this case, we may regard the observations as constraints and set up an optimizing function of the parameters alone.
3. The total number of observations is about the same as the number of parameters. In this case, we may try to find or establish enough constraints so that the total number of constraints and observations is much greater than the number of parameters.

Kinds of Relationships

Linear Relationships

The two specific models we have discussed are fairly simple in four respects:

1. They assume a linear combination of observed values.
2. They assume the parameters to be independent of one another.
3. The typical loss function is a sum of squared residuals so that it is at most of second degree in the parameters.
4. Because of (3), the partial derivatives of the loss function with respect to the parameters are of first degree in the parameters. Therefore, the partial derivatives are linear in the parameters. This linearity makes for a straightforward solution for the parameters.

Nonlinear Relationships

Some models, however, are not linear in the parameters. Among the best known of these are the various factor analytic models. It is also probable that special cases of a comprehensive expectancy theory model may turn out to be nonlinear in the parameters of the model although, as mentioned earlier, most of the current models have attempted to get along without the introduction of parameters. The parameters of a model are typically the most interesting and scientifically significant features of the model and their absence in expectancy theory marks it as a scientifically primitive example of model building.

In any case, when a model postulates nonlinear relationships among the parameters, the solutions for them can become computationally involved, requiring iterative procedures. An important part of much multivariate analysis activity consists of developing solutions for the parameters when they are nonlinear in the mathematical function of the model or when the loss function is not a simple quadratic form in the parameters.

Summary of Multivariate Analysis Systems

We shall now briefly summarize a multivariate analysis system.

The Structural Component

A system consists of its structural elements and the functional relationships of the structural elements.

1. A structure consists of a set of at least two modalities, one of which is an entity modality and another an attribute modality.
2. A modality consists of one or more dimensions.
3. The dimensions of some modalities consist of two or more levels.
 - a. Dimensions may consist of the levels of a dimension.
 - b. Dimensions may be manifest or latent.
 - c. Dimensions may be predictor or criterion.
 - d. Dimensions may be manipulable or non-manipulable.
 - e. Dimensions may be endogenous or exogenous.
 - f. Dimensions can be generated from the products of two or more other dimensions.
4. Levels may be binary, n-ary, or continuous.
5. A variable is a dimension with its associated levels.

The Functional Component

The functional part of a multivariate analysis system begins with a mathematical model, which consists of variables and parameters. We begin with special cases of the model, one corresponding to each dimension of the entity modality. These are sometimes called observation equations. The group of entity dimensions is sometimes called the sample.

A loss function is constructed from the given values of the variables corresponding to each of the dimensions of the attribute and entity modalities and the unknown parameters. One or more constraining functions of the parameter may be specified. A total function is set up that is the sum of the loss function and the products of unknown constants times the constraining functions. The partial derivatives of the total function are taken with respect to each of the parameters of the model and equated to zero, thus yielding as many equations as we have parameters. These equations are functions of both the unknown parameters and the unknown Lagrangian multipliers corresponding to the constraining equations. The unknown Lagrangian multipliers may be eliminated from the derived equations through the use of the constraining equations. The unknown parameters are solved for by means of the derived equations from which the Lagrangian multipliers have been eliminated.

The value of the loss function or optimizing function is calculated by substituting in it the solved for values of the parameters. Estimates of the appropriateness of the model are obtained as a function of the optimizing function, the number of observations, the number of constraints, and the number of parameters. If the total number of observation and constraining equations is less than the number of parameters, a test of the model is not possible. In such cases, it is sometimes possible to formulate an optimizing function

of the parameters alone and thus introduce enough degrees of freedom in the system to make possible tests of significance for the model.

Optimal Parameter Estimation

The problem of estimating parameters from a sample that will hold up for subsequent samples is one of the most fundamental in science. It arises whenever the number of observations is not extremely large compared to the number of parameters in the model. Unfortunately, such situations are the rule rather than the exception. Aside from the work of Horst (1970), Leiman (1951), and Burket (1964), little has been done to attack the problem. The ridge regression approach, which inflates the diagonal elements of the covariance matrix, appears to be an effort in this direction but so far its contribution is not conclusive. The traditional cross-validated approaches are crude and constitute much more of a recognition of the problem than its solution.

Examples of Multivariate Modeling

In the Vroom (1964) activity choice model, an experimental design can yield two matrices of experimental measures for the same group of persons and the same set of activities. These two matrices are assumed to yield the same values, except for person, origin, and scaling parameters to be solved for. Appendices A through H suggest models that vary in the constraints imposed for the scaling and origin parameters.

A MORE ADEQUATE EXPECTANCY THEORY MODEL

We are now ready to begin an in-depth consideration of a more adequate expectancy theory model. We shall attempt to bring it into the mainstream of multivariate analysis systems where these can accommodate the necessary concepts of the comprehensive model. Where this is not possible or obvious, we shall attempt to expand the available models as the need develops.

The Environments

First we shall try to lay the groundwork for our formulations in terms of organisms (persons) and their environments. To accommodate the key concepts of current expectancy theory, we recognize five types of environment: organic, exterior, ambient, focal, and contingent, which are described below:

1. The organic environment may also be called the "internal environment" of the organism and may appear to be a contradiction in terms. Here we include any internal physiological states or processes, chemical, physical, or whatnot that influence the behavior of the organism with respect to either his internal or external environment.

2. The exterior environment refers to that part of the environment that is independent of and outside the job. It may include the home, the spouse, children, the neighborhood in which a person lives, neighbors, the community, the church, and any other aspects of a person's world that may directly or indirectly influence his behavior on the job, except the job environment itself.

3. The ambient environment includes that part of the total job environment that may directly or indirectly influence his behavior on the job. It may include that part of the entire physical plant of the company to which a person is exposed, the buildings, the

grounds, the offices or shops and their equipment, furnishings, etc. It would also include the other workers in the company with whom he comes in contact.

4. The focal environment includes that part of the job environment that a person is supposed to modify in some way. It is his or her specific job. He or she does something to something on an assembly line, punches cards according to some written documents, types a letter with a typewriter on a piece of paper according to certain audio signals received from a dictating machine, assembles a part of a radio from separate pieces on the work table, or sews pieces of cloth together to make a garment. The focal environment is that part of the environment that he or she is assigned to change in some prespecified way. The focal environment plays a most crucial role in the application of expectancy theory to organizational behavior.

5. The contingent environment is that part of either the internal, ambient, or focal environment that may influence the way in which a person modifies the focal environment. He or she has a headache or feeling of euphoria that inhibits or facilitates the modification of the focal environment in a prespecified manner. The records from which the cards are punched are illegible; some of the parts that are to be assembled don't fit together; the sewing machine keeps breaking thread and slows down the sewing of garments; a jack hammer is breaking up concrete outside the window; fellow workers are laughing or gossiping at a nearby desk.

It is true that the lines cannot always be clearly drawn between these five types of environments but they all have appeared either implicitly or explicitly in the expectancy theory literature. Perhaps the least emphasized of the five is the external environment, even though it may play an important role in organizational behavior. The most emphasized in the literature is the focal environment. This concept is directly involved in the concepts of "performance" and "task goals," which play such an important part in the current models of expectancy theory.

Now for a specified time or state of each type of environment we may assume an attribute modality with a number of dimensions and levels of each dimension such that the environment may be adequately characterized by them. This notion of the possibility of adequately specifying the dimensions and their levels for each type of environment becomes more crucial as we attempt to develop the structural elements and their functional relationships for a more comprehensive expectancy theory model. However, before we attempt to develop further the notion of the structural elements of the environment, we must return again to the organism or person.

The Organism

The organism has an internal environmental system and an operant or responding system. We have already discussed briefly the internal environmental system, which may be identified with the internal stimuli or sensory phenomena of the organism. The operant system may be identified with the motor activities of the organism as produced by striated and smooth muscle and glandular activities.

Each system may be thought of as having an attribute modality with dimensions, each of which has levels. Thus, both the organic system and the environmental system have an attribute modality with dimensions and their associated levels. The operant system of an organism may modify the state of an environment. A level of a dimension of an environmental modality may activate the operant system of the organism. The valence of such a level for a person is the extent to which it activates the operant system to maintain or increase or reduce that level for the organism.

The Interactive vs. the Expectancy System

Let us see now how the concepts of organismic and environmental modalities with their dimensions and associated levels may be articulated with the traditional expectancy theory concepts--personal attributes, jobs, effort, performance, outcomes, need satisfaction, valence, and expectancy. We shall call the organismic-environmental system the "interactive system," and the traditional expectancy theory model the "expectancy system."

The Organismic Modality

In both systems, we begin with the person or organismic modality whose dimensions are the persons in the system. In both systems, we assume a person characteristic modality whose dimensions may include a wide variety of aptitude, proficiency, personality, biographical, and other subgroups of dimensions, with levels specified for each dimension.

The Attribute Modality

In the interactive system, we segregate within the attribute modality the two subgroups of dimensions that constitute the operant and internal environment modalities. These modalities are not explicitly specified in the expectancy system.

The Job Environment Modality

In both systems, we have a job environment modality. In neither system do we define the term "job" precisely. A job may be in a department within a company. It may be a specific kind of job, irrespective of what company or organization it is in, such as a card punch operator, secretary, typist, etc. In general, however, both systems attempt to distinguish between job environment and job activity.

The job environment may, in both systems, be conceived as a set of two modalities: (1) the job or environmental modality and (2) a jobs characteristic modality with its dimensions and levels. This conceptualization is compatible with the interactive system and with Vroom's (1964) job choice model in the expectancy system. In the interactive system, the job is referred to as the ambient environment.

The Focal Environment Modality

Both systems involve a focal environment modality, which is explicit in the interactive system. It is implicit in the "performance" and "task goal" concepts but no serious attempt is made to analyze these concepts in terms of the operant modification of the focal environment. It is at this point that the expectancy system begins to run into semantic and conceptual difficulties.

The interactive system assumes a preoperant specified state of the focal environment modality defined in terms of characteristic dimensions and associated levels. It also assumes a prespecified postoperant state of the focal environment defined in terms of its characteristic dimensions and associated levels. Neither of these assumptions is made explicit nor more than remotely implied in the expectancy model.

Both models must assume a restructuring of the focal environment by means of the operant modality of the organism. This operant modality may be defined in terms of its dimensions and associated levels. In the expectancy system, this modality is usually

regarded as consisting of only one dimension and is called "effort." The literature on expectancy theory does not emphasize the unidimensionality of effort but the usage clearly carries this assumption. Whether or not an extension to more than one dimension would be fruitful remains to be tested. Certainly there can be both actual muscular exertion and mental or neural effort. Traditionally also the levels of this unidimensional modality have been restricted to two, or at most three, levels. This restriction, like others in expectancy theory models, has been dictated in part by limitations in the ability of subjects to make reliable discriminatory judgments with more than a few levels.

The Contingency Environment Modality

Both systems may assume a contingency environment that may influence the effect of effort on the extent to which the operant activity can restructure the focal environment to prespecified standards. This contingency environmental modality has characteristic dimensions with associated levels. The extent to which the contingent environment can impede or facilitate the operant activity in restructuring the focal environment may be evaluated or estimated. Such estimates have been incorporated into expectancy theory models (Nebeker & Moy, 1976).

The Restructured Focal Environment Modality

Both systems assume a restructured focal environment that can be compared with prespecified restructured standards. This is called job "performance" or "task goal accomplishment." In the interactive model, we may assume a multidimensionality for the focal environment characteristics modality. Job performance may therefore be a multidimensionality modality. At a minimum, we may have a speed or volume and an accuracy or quality dimension. Each of these may have two or more levels. The expectancy system has usually restricted performance to one dimension. Studies by Nebeker and Moy (1976) have used the speed or volume concept and have restricted it to seven levels.

It should be noted that both effort and performance have been used as criterion dimensions and that the levels of these have been expanded to serve as dimensions in expectancy choice models. We shall have more to say of this treatment of levels as subgroups of dimensions with binary levels. The procedure plays an important role in expectancy theory models, whereas it has been little used in traditional multivariate analysis models. This added complication introduces problems into the expectancy theory models that are not ordinarily encountered in the traditional models.

The Outcome Modality

Although the outcome modality is assumed in both the interactive and the expectancy systems, there is considerable ambiguity and variation in the usage of this term in the expectancy theory literature. Sometimes "performance" or "goal task accomplishment" is regarded as an outcome modality; sometimes even "need satisfaction" is called a second-level outcome. Typically, however, in expectancy theory research, the manifest dimensions of the outcome modality are defined by a list of statements describing events that may follow specified effort or performance levels and that may have value (valence) for a person. Such a list constitutes the definition of the outcome modality and sometimes provides for both intrinsic and extrinsic outcomes.

The outcome modality in the interactive system may be characterized by the fact that chronologically the events represented by the dimensions of the modality occur after the focal environment has been restructured. These events may constitute a restructuring

of the internal environment of the person or a restructuring of the exterior, the ambient, or the contingent environment. In either case, the dimensions of the characteristic modalities of these environments may assume new levels. It is to these altered levels that the concepts of valence are attached.

In the expectancy system, the restructured ambient or contingent environment is referred to as "extrinsic" while the internal environment is referred to as "intrinsic." An important distinction in both systems is between those events that are initiated by forces other than the person himself and those which are not. The literature is a little hazy on this latter alternative. Some writers say that the distinction is between rewards conferred by the person and those conferred by other persons or instrumentalities.

The Needs Modality

Finally, both systems assume a "needs" modality but, as suggested earlier, it is not the needs but their satisfaction that constitutes the event. Presumably, this is the final event in the sequence of effort, performance, outcome, and need satisfaction. The actual research reported in expectancy theory literature does not appear to have done much with the concept. In the interactive model, it is the internal environment that is involved. Whether or not it is necessary to postulate a component of the internal environment that is distinct from that that is restructured by "effort" or "performance" is not yet clear. Perhaps a distinction between a relatively stable genetic physiological structure and a more transient condition of the organism will be useful. The expectancy theorists do not appear to have systematically considered this question.

Summary of Modalities

We shall now summarize the chain of modalities for the interactive system. We begin with an external environment consisting of four environmental modalities: (1) the exterior environment, (2) the ambient environment, (3) the contingent environment, and (4) the focal environment. The last three may be called the activity-environmental set.

Next we introduce the organismic set of modalities consisting of the entity set of person and attribute modalities. The dimensions of the attribute modality consist of subgroups of operant dimensions, stable dimensions, and transient internal environmental dimensions. The third subgroup is identified with what is sometimes referred to as "intrinsic rewards"; and the second, with "needs satisfaction."

Initially, the person is introduced to the activity-environmental set of environmental modalities. In this set, the focal environmental modality has an initial state with respect to the levels assigned to its characteristic dimensions. He is also provided with specifications for a restructured focal environment.

The modality of operant dimensions of the organism operates on the focal environment to restructure it according to the specifications for the restructured focal environment. This complex of operations is called "effort." The effort modality may be multidimensional and the levels of the dimensions may be more than two.

The level values of the dimensions of the restructured focal environment may be compared with the specifications for this restructured focal environment. The extent of agreement (however measured) of the restructured focal environment with the specifications for it may be called the performance level. Or, if the performance modality has more than one dimension, we may speak of the performance levels attained.

The level or levels of performance attained may be regarded as functions of (1) the status of the operant dimensions of the organism, (2) the degree and relevance of the effort dimensions exerted, (3) the status of the contingent environment dimensions, and (4) the difficulty of the restructuring processes. Of these four factors, perhaps the second requires most clarification. The degree of effort exerted is perhaps the most elusive concept and may come closer to the concept of "motivation," as it is used in expectancy theory literature, than any of the other factors related to performance. The relevance of the effort level exerted has not been well articulated into expectancy theory formulations. The notion of "understanding of the task to be accomplished" is sometimes mentioned as an important aspect of job performance. Perhaps this comes close to relevance of effort.

The difficulty of the restructuring process is sometimes implied as a factor in performance but the concept does not appear to have been systematically incorporated into schematic representations of expectancy theory. Eventually, however, it must be, as it is fundamental to psychological measurement theory and technology. An attempt has been made by Ilgen, Pritchard, Dugoni, Matte, and Nebeker (1980) to vary difficulty by varying the numerical coding of performance levels.

Summary of the Interactive System

We see then that the interactive system must begin with an environment into which an organismic system is introduced that exerts effort on the focal part of the environmental system to produce a performance or product. This sequence is followed by an outcomes system that consists of modifications of one or more components of the environmental system. One of the components may be the internal environment component of the system. Those modified components of the outcomes system exclusive of the internal environment may be followed by a modification of the stable component of the internal environment modality; namely, the "satisfaction of basic needs."

Valence and Expectancy

We have now indicated a reorientation of the expectancy theory concepts of ability, effort, performance, outcome, and needs around the interactive system whose major components are the environment and the organism. We are now ready to resume in greater depth our discussion of the concepts of valence and expectancy in the light of this reorientation.

Earlier, we attempted to define the valence of an event in terms of the organism's reactions to a stimulus or environmental situation. The stimulus situation was defined as positively valent if the organism's responses tended to maintain or enhance it and negatively valent if they tended to reduce or eliminate it. This definition will now be modified somewhat. Some examples could be cited of stimulus situations that a person does not like and that could be said to be negatively valent, yet he might be powerless to reduce or eliminate the situation. For example, he may have a job that he loathes; yet, for all practical purposes, he may be powerless to remove himself from the job environment. There may be high overall unemployment, it may be necessary for this person to work, and there may be no other jobs available that he would like better.

Valence in Terms of Alternatives

This example suggests that, theoretically at least, valence should be defined in terms of available environmental alternatives. We may assume a set of hypothetical scaled alternative environmental situations. These would range from very high negative to high

positive. We assume that any and all of these would be available to the organism. For any given actual environmental situation, the person could compare it with each of the hypothetical situations. For each of the hypothetical situations, he would indicate whether he would prefer it or the actual situation or whether he was indifferent. The valence for him of the actual situation would be an average or some other central tendency of the scale values of the hypothetical situations for which, if compared with the actual situation, he would express indifference. This operational definition of the valence of an environmental situation can thus include the "approach" and "avoidance" type of definition, while at the same time also accommodating those situations where actual approach or avoidance is physically or practically impossible.

We note that our definition avoids the philosophically and operationally unfortunate terms of "pleasure-pain," "reward-punishment," and other hedonistic designations. Even the unhappy but persistent designations of "positive and negative reinforcement" have been circumvented, although perhaps not liquidated. Furthermore, it must be emphasized that the operational definition we have suggested for valence does not necessarily constitute a recommendation for an experimental procedure for determining the valence of an environmental situation. However, the definition may suggest possibilities for eliciting valence levels from persons for specified environmental situations.

Valence may then be defined as the tendency to react so as to approach, enhance, or maintain, or to avoid, reduce, or eliminate a stimulus situation, on the assumption that all these alternatives are available to the organism. The proviso is an important modification of the traditional approach-avoidance definition. It is also important to emphasize that valence refers to any component of the relevant stimulus environment of the organism, even though relevancy may be difficult to determine in some cases.

Concomitance of Valence and Expectancy

The concept of expectancy, as we have seen earlier, has chronological or historical implications that valence does not have except indirectly. For the most part, as we shall see, any environmental situation that has valence for a person also may have some degree of expectancy of occurrence on his part. A person may have some degree of expectation between zero and one that a specified environmental situation can or will occur. He may also be able to say how desirable (valent) the situation is for him. He may have some degree of expectation that a given environmental situation will be followed by another. And, if we extend the environmental situation to include a chronological sequence of environmental situations, a person may be able to express how valent this sequence of situations is for him.

Expectancy and the Operant Event

There is one type of event or situation that may have an expectation associated with it but which may not properly be regarded as valent in and of itself. This is the operant event produced by the organism. A person may have an expectation that he can or will effect certain muscular or glandular activities, whether or not these will modify the state of any part of his external environment. Could the activity itself also be valent for the person in the sense that we have defined the term? One may argue that the flexing of muscles or secretion of adrenalin can also be more or less pleasant or unpleasant. But one could equally well maintain that it is the internal environment produced by the muscular or glandular activity that has valent potential rather than the activity itself. For purposes of model formulation, it may not be essential to discriminate between the activity that has an expectation of occurrence and the valent internal environment associated with the activity. One could, of course, conceive of the expectation of an

organismic activity as well as the expectation that a specified valent internal environment will be associated with the activity. However, until the need arises for making an exception, we shall assume that every valent situation may also have an expectation of occurrence and that every situation that has an expectation may also be valent.

Expectancy and Experience

We have already attempted to give an operational interpretation of the expectancy concept. This interpretation has been called historical or chronological because it is based on the concept of a recall of past events. The expectancy of the occurrence of an event or sequence of events is based directly or indirectly on the recalled frequency of the occurrence of such events or sequences in the past under comparable conditions.

To stretch the concept a bit to make it more general, suppose a person is told that, if he reaches a specified level of productivity, he will get a raise of \$50 per month. He has never in the past achieved this rate of productivity so he has no direct past experience on which to base his expectation that he will get the raise if he reaches the specified level. However, he has been told by fellow workers that the company, for whatever reasons, fails to follow through about half the time on announced promises to its employees. Therefore, he may express his expectation as 50 percent that, if he achieves the specified production level, he will get a raise of \$50 per month. His expectancy is based on what he has been told about the proportion of times the company has followed through on its commitments.

Thus, probably any expression of the expectancy of occurrence of an event can be traced directly or indirectly to the alleged or reported proportion of times that a given event has occurred under a specified pattern of circumstances. How accurate the observations or reports may be is quite a separate question. The important point is that the expectation is the person's estimate or perception of the probability. It is this perception that is the crucial determinant in his behavior.

Expectancy and Certainty

The certainty with which a person fixes on a particular probability figure can be significant. Although the distinction between expectancy and certainty is sometimes made in the expectancy theory literature, it does not appear to have been introduced into experimental investigations. There is a big difference between saying, "I am 60 percent certain the chances are close to 50 percent that A will happen," and "I am only about 20 percent certain that the chance of A happening is 50 percent." The two figures of 50 percent are presumably based on two different sets of data available, directly, or indirectly, to the subject. The other two percentages represent his confidence in each set of data available to him. The confidence is based on his estimate of the adequacy of the data.

It would be entirely possible to include this second type of estimate in a comprehensive expectancy theory model. There has been some experimentation with a similar concept in psychological measurement methodology. For example, in a multiple-choice type of item, the subject, in addition to checking what he thinks is the right answer, may also express his degree of confidence in his choice. It is quite possible, however, that the introduction of this additional concept into the expectancy theory model would have to be at the expense of suppressing or oversimplifying some of the other concepts. The major consideration is to keep the model simple enough so that persons can make the discriminations called for with an acceptable degree of reliability.

EXPECTANCY THEORY AS A MULTIVARIATE ANALYSIS SYSTEM

We are now ready to begin our discussion of an expectancy theory system as a special case of a multivariate analysis system. Earlier, we have presented the latter as consisting of two major aspects: (1) the structural elements of the system and (2) the functional relationships among those elements. We have defined (1) a set as consisting of two or more modalities, (2) a modality as consisting of one or more dimensions, and (3) some dimensions as having two or more levels. We have seen how additional dimensions can be generated both from the products of two or more dimensions and from the levels of dimensions.

We have already discussed the reduction of the major expectancy theory concepts to those involving the interaction of the organism with various types and sequences of environments. We shall take the position that any concept or construct in expectancy theory that cannot be translated directly into terms of organism, environment, and their interaction is suspect as a semantic evasion whose utility should be seriously challenged.

The Abstract Model

We shall have a number of problems to consider that arise in current expectancy theory models. These can best be discussed if we first outline an abstract model that, hopefully, will be more adequate than current models to accommodate the behavior phenomena that current expectancy theory attempts to cope with.

We begin with a set of modalities, some of which are logically or chronologically ordered. We use the terms prior and subsequent to indicated the logical or chronological sequence of any two modalities.

Two special types of modalities have dimensions that have only one level. These are called assignee and assignor modalities. The assignee modality is one to whose dimensions levels of dimensions of other modalities are assigned. The assignor modality is one whose dimensions assign levels of dimensions of other modalities to dimensions of the assignee modality.

The dimensions of an assignee modality may be people. For exmaple, the assignee modality may be persons, each of whom constitutes a dimension of the people modality. The assignor may be raters, or evaluators, each of whom constitutes a dimension of the rater modality. Each rater may assign a level of a dimension of a third modality to some or all of the people dimensions of the assignee modality. The third modality may be personality traits, each one of which constitutes a dimension of the personality modality. The levels of the traits dimensions may go from zero to ten. A given assignor (rater) might assign a level of five on "aggressiveness" to a given assignee (person). It is important to note, however, that the dimensions of only the personality modality have associated levels. A dimension of the assignee level does not have associated levels; he or she is either included in a specified data set or not. Likewise, a dimension of the assignor modality, a rater, is either included in the data set or not. The person-rater-trait data set exemplifies what Tucker (1963) has called the three-mode data matrix. From this model, he has developed his three-mode factor analysis procedures.

The assignee and assignor modalities may be the same. This means that assignments of levels of dimensions of other modalities may be self-assigned.

A Notational System

At this time, it may be helpful to introduce a notational system as follows:

| | |
|-------------------------|---|
| (a) | is the assignee modality. |
| (c) | is the assignor modality. |
| (a,b) | is dimension (b) of modality (a). |
| (c,d) | is dimension (d) of modality (c). |
| (f) | is a modality subsequent to modalities (a) and (c). |
| (i) | is a modality subsequent to modality (f). |
| (f,g) | is dimension (g) of modality (f). |
| (i,j) | is dimension (j) of modality (i). |
| (f,g,h) | is level (h) of dimension (g) of modality (f). |
| (i,j,k) | is level (k) of dimension (j) of modality (i). |
| n(a,b) | is the number of dimensions (b) of modality (a). |
| n(c,d) | is the number of dimensions (d) of modality (c). |
| n(f) | is the total number of modalities subsequent to modalities (a) and (c). |
| n(f,g) | is the number of dimensions (g) of modality (f). |
| n(i,j) | is the number of dimensions (j) of modality (i). |
| n(f,g,h) | is the number of levels (h) of dimension (g) of modality (f). |
| n(i,j,k) | is the number of levels (k) of dimension (j) of modality (i). |
| P(a,b,:c,d:f,g,h) | is the expectation level assigned to dimension (b) of modality (a) by dimension (d) of modality (c) that level (h) of dimension (g) of modality (f) <u>can</u> or <u>will</u> occur. |
| P(a,b,:c,d,:i,j,k) | is the expectation level assigned to dimension (b) of modality (a) by dimension (d) of modality (c) that level (k) of dimension (j) of modality (i) <u>will</u> or <u>can</u> occur. |
| V(a,b:c,d:f,g,h) | is the valence level assigned to dimension (b) of modality (a) by dimension (d) of modality (c) for level (h) of dimension (g) of modality (f). |
| V(a,b:c,d:i,j,k) | is the valence level assigned to dimension (b) of modality (a) by dimension (d) of modality (c) for level (k) of dimension (j) of modality (i). |
| P(a,b,:c,d:f,g,h:i,j,k) | is the expectation level assigned to dimension (b) of modality (a) by dimension (d) of modality (c) that level (h) of dimension (g) of modality (f) will be followed by level (k) of dimension (j) of modality (i). |
| V(a,b:c,d:f,g,h:i,j,k) | is the valence level assigned to dimension (b) of modality (a) by dimension (d) of modality (c) for level (h) of dimension (g) of modality (f) when followed by level (k) of dimension (j) of modality (i). |

The notation we have just defined appears inordinately complicated. However, it merely makes explicit the concepts that have been obstreperously slithering about in the luxuriant semantic jungle of current expectancy theory models.

The Instrumentality Concept

Concepts corresponding to each of the above defined symbols may be identified in actual reported experiments or model formulations. We shall identify some of the less obvious ones but first we must put to rest one of the most troublesome concepts currently used in expectancy theory models. This is the instrumentality concept.

Two features of this concept have been defined to distinguish it from expectancy. First, it is said to be an outcome-to-outcome relationship while expectancy is said to be an action-to-outcome relationship. Second, it is said that expectancy is measured in terms of probability ranging from zero to one while instrumentality is measured in terms of a correlation ranging from minus one to plus one.

The Modality of Instrumentality

Both of these measures are subjectively arrived at by the subject. It is not difficult to believe that a person might be willing and able to express his expectation that one event (any event) will be followed by another event (any other event). So why distinguish between types of events as a basis for specifying the type of sequential relationship? As we have pointed out earlier, there seems to be no pressing need for separating the motor activity of the organism from the internal environment associated with it. Therefore we submit that action-outcome and outcome-outcome distinctions do not justify a type of relationship distinction and that, for the sake of both parsimony and clarity, a single concept (expectancy) is preferred.

Instrumentality as a Correlation

Next let us examine the concept of instrumentality as the correlation between two outcomes. Although the point is not always made explicit, the standard formulation implies that there is one group of outcomes closely clustered in time and a second temporally-clustered group of outcomes chronologically subsequent to the first group. Each outcome of the first group is said to have a set of instrumentalities corresponding to each of the second group. These instrumentalities are expressed as correlation coefficients; herein lies the feet of clay of the instrumentality concept, as we shall now see.

We let O_g be the g th outcome (dimension) of the prior group (modality) of outcomes and O_j be the j th outcome (dimension) of the subsequent group (modality) of outcomes. Then, according to the instrumentality concept, I_{gj} is the instrumentality of O_g for evoking O_j . According to instrumentality doctrine, I_{gj} is a correlation coefficient. If indeed it is a correlation coefficient, there seems little excuse for the notation I rather than the well established r_{gj} . However, having committed ourselves to this well-nigh inescapable substitution, we have introduced some crucial implications which we must follow through to the bitter end. And this we shall now proceed to do.

A correlation coefficient implies two variables that, in the case under consideration, we shall call a g and a j variable. We have previously defined a variable as a dimension with its associated levels. Here we introduce a concept that, simple as it seems, has not been emphasized by most expectancy theory model builders. It is one for which Nebeker and Moy (1976) recognized the need--the simple notion that every dimension of an

outcome modality must have at least two levels. This is the simplest case, in which an outcome occurs or does not occur. Without loss of generality, the first level outcome may be assigned the value of 0 for nonoccurrence and the second level, 1 for occurrence.

How then do we calculate a correlation coefficient from two such binary variables?

We let

P_g be the expectation that outcome g will occur,

P_j be the expectation that outcome j will occur, whether or not g occurs,

and

P_{gj} be the expectation that the occurrence of outcome j will follow the occurrence of outcome g .

Then the Pearson product moment coefficient of correlation for binary variables is well known to be

$$r_{gj} = \frac{(P_{gj} - P_g P_j)}{((P_g - P_g^2)(P_j - P_j^2))^{1/2}} \quad (62)$$

It is well known that this coefficient cannot be plus one unless $P_g = P_j = P_{gj}$. Thus, if the expectation of outcome g is greater or less than that of outcome j , the instrumentality of outcome g for outcome j cannot be unity, no matter how high the expectation that outcome j will follow outcome g . Other limiting cases of the so-called phi coefficient are discussed by Horst (1966) for both positive and negative values of phi.

Apologists for the instrumentality concept could protest, however, that they are not thinking of a phi coefficient but, rather, of a tetrachoric coefficient or some less mathematically elegant type of contingency coefficient. Such protests, however, could not be taken seriously for at least two reasons. First, it is presumptuous enough to expect unsophisticated subjects to (1) estimate probability values for P_g , P_j , and P_{gj} in one fell swoop and (2) convert them to a phi coefficient with a flash of their neuronics computers. Second, it is preposterous to expect them to recognize and select from available binary-type correlation coefficients and then come up with the required estimates and calculations. One might be reluctant to abandon the concept of instrumentalities, however presumptuous the possibility of estimating them might be, if they could be put to good use once obtained. But the valence model of Vroom (1964) is given by

$$V_j = f \sum_{k=1}^n (V_k I_{jk}) \quad (63)$$

where

V_j = the valence of outcome j ,

I_{jk} = the cognized instrumentality of outcome j for the attainment of outcome k ,

V_k = the valence of outcome k ,

and

n = the number of outcomes.

(This notation is not to be confused with that which we have already defined.)

Here, presumably, V_j is the valence of level 1 of outcome (dimension) j of a prior modality of outcomes, and V_k is the valence of outcome (dimension) k of a subsequent modality of outcomes. This formulation as presented cannot be regarded as meaningful, among other reasons, because no assumptions are specified about the metrics of V_j or the V_k nor about the interrelationships among the V_k . We shall have more to say about this issue later. For the moment, we turn our attention to the I_{jk} .

Even assuming the I_{jk} can be interpreted and estimated as correlation coefficients, one must search hard to find an example where a weighted sum of variables can be reasonably interpreted when the weights are zero-order correlation coefficients. One example comes to mind. Suppose the subsequent group of variables is regarded as standardized predictor variables that are mutually orthogonal, and each I_{jk} is the validity coefficient of the predictor k with the standardized criterion variable j . Then the weighting of the predictor variables by the corresponding validity coefficients would yield a composite that would be the best estimate of the criterion variable in the least square sense.

The Levels of Dimension

Let us now return to the two concepts of "action-to-outcome" and "outcome-to-outcome" and remind ourselves that they can involve no more than organismic or environmental events. The subsequent outcome, whether following an action or some other outcome, has already been recognized by Nebeker and Moy (1976) as at least a two-level dimension of an outcome modality. This characteristic of an outcome does not appear to have been emphasized by Vroom (1964) or most of his followers. However, we can now speak of the valence of the first level (zero) of an outcome; that is, the valence of the nonoccurrence of the outcome, as well as the valence of its occurrence. Simplistic as it may be to restrict to two the levels of a dimension of an outcome modality, this is at least a step in the right direction.

Two-level Dimensions

With this slight concession to reality, we can also speak of the expectancy that level 1 of the outcome will follow an action or a prior outcome. We can also speak of the expectancy that level zero of the outcome (nonoccurrence) will follow an action or a prior outcome. Logically, the sum of these two expectancies should equal one. It is, of course, possible that, in some experimental settings, subjects are not consistent or logical in making such judgments, but this is another matter that does not concern us here. We shall assume that with suitable procedures the two will add up to one.

Multi-level Dimensions

There is no logical reason to restrict the number of levels of the subsequent outcome dimension to two. As a matter of fact, most outcomes might more realistically be assigned substantially more than two levels. Then we could speak of the valence of each

level and also the expectancy that each of the levels would follow a given action or prior outcome. The total of these expectancies would be constrained to add up to one.

However, this proliferation of the number of levels of an outcome dimension suggests that we examine more closely the concept of the prior action or outcome. Suppose we consider the concepts of effort and performance that we have already discussed and translated into organismic-environmental terminology. We have seen that, while each could be regarded as a multidimensional modality, they have typically been discussed in the literature as unidimensional, and with each of these modalities two or more levels have been associated. Thus, one may speak of the expectancy that a given level of either of these modalities will be followed by a given level of a subsequent outcome. In the limiting case, one could speak of the expectancy that either zero effort or zero performance would be followed by a specified level of a subsequent outcome.

Expectancies for Pairs of Dimensions

We are then led to suggest that even for the most primitive of expectancy theory models, one may not properly speak of the expectancy that an act or prior outcome will be followed by a subsequent outcome. Rather, we must speak of the expectancy that a given level of a given act or prior outcome will be followed by a given level of a given subsequent outcome. In the simplest acceptable model, we identify the following four expectancies with reference to a prior and a subsequent outcome where we include an action also as a prior outcome:

1. The joint expectancy that the prior outcome does not occur and the subsequent outcome will not occur, which we call P_{00} .
2. The joint expectancy that the prior outcome does not occur and the subsequent outcome will occur, which we call P_{01} .
3. The joint expectancy that the prior outcome does occur and the subsequent outcome will not occur, which we call P_{10} .
4. The joint expectancy that the prior outcome does occur and the subsequent outcome will occur, which we call P_{gj} .

We can see now that the expectancy that the prior outcome will occur is the sum of the expectancies (3) and (4) above. This we call P_g . Also, the sum of (2) and (4) above is the expectancy that the subsequent outcome will occur. This we call P_j . We have now come full circle from the instrumentality concept that we defined in line with Vroom's interpretation as a phi coefficient. The expectations P_{gj} , P_g , and P_j are precisely those we used in equation (62) to calculate the correlation coefficient.

The Nebeker-Moy Model

Before we can decide to dispense with the instrumentality concept, we must examine more closely the concept of expectancy in connection with the levels of dimensions of prior and subsequent modalities.

The Algebra of the Model

We begin by reproducing an equation presented by Nebeker and Moy (1976). This is

$$F_i = \sum_{j=1}^n (E_i P_{ij} V_j + (1 - E_i P_{ij}) \bar{V}_j) \quad (64)$$

where

F_i = force to perform at level i ,

E_i = the individual's expectancy that he could perform at level i ,

P_{ij} = the perceived likelihood that performance at level i will result in outcome j ,

V_j = the valence of obtaining outcome j , and

\bar{V}_j = the valence of not obtaining outcome j .

Let us see now how we may translate the formulation in equation (64) into an instrumentality estimate. First, to simplify the development let us write:

$${}_jF_i = E_i P_{ij} V_j + (1 - E_i P_{ij}) \bar{V}_j \quad (65)$$

and

$$F_i = \sum_{j=1}^n {}_jF_i \quad (66)$$

Presumably, equation (65) gives the contribution of outcome dimension j to F_i . However, in this equation we have in effect recognized that the outcome dimension has two levels and we have assigned the values of 1 and 0 to them. Thus, V_j is level 1 of the outcome and \bar{V}_j is level 0. Let us next drop the subscript j and let

V_1 = the valence of level 1 of the outcome, and

V_0 = the valence of level 0 of the outcome.

Assume that there are two levels of performance, 1 and 0, and let

ρ_{00} = the expectation that performance level 0 and outcome level 0 will occur,

ρ_{01} = the expectation that performance level 0 and outcome level 1 will occur,

ρ_{10} = the expectation that performance level 1 and outcome level 0 will occur,

ρ_{11} = the expectation that both performance level 1 and outcome level 1 will occur,

ρ_0 = the expectation that performance level 0 will or can occur,

and

ρ_1 = the expectation that performance level 1 will or can occur.

Then, according to the definitions,

$$\rho_{00} + \rho_{01} = \rho_0 \quad (67)$$

$$\rho_{10} + \rho_{11} = \rho_1 \quad (68)$$

$$\rho_{00} + \rho_{01} + \rho_{10} + \rho_{11} = 1. \quad (69)$$

Suppose now we let

$$\rho_{00} / \rho_0 = P_{00} \quad (70)$$

$$\rho_{01} / \rho_0 = P_{01} \quad (71)$$

$$\rho_{10} / \rho_1 = P_{10} \quad (72)$$

$$\rho_{11} / \rho_1 = P_{11}. \quad (73)$$

From (67), (70), and (71)

$$P_{00} + P_{01} = 1. \quad (74)$$

From (68), (72), and (73)

$$P_{10} + P_{11} = 1. \quad (75)$$

Now in the notation of equation (65)

$$E_0 = \rho_0 \quad (76)$$

$$E_1 = \rho_1. \quad (77)$$

We let

$$V_j = V_1 \quad (78)$$

$$\bar{V}_j = V_0 \quad (79)$$

$$P_{i1} = P_{ij} \quad (80)$$

$$P_{i0} = 1 - P_{i1}. \quad (81)$$

Substituting (74) through (81) into (65),

$${}_jF_0 = \rho_0 P_{01} V_1 + (1 - \rho_0 P_{01}) V_0 \quad (82)$$

$${}_jF_1 = \rho_1 P_{11} V_1 + (1 - \rho_1 P_{11}) V_0. \quad (83)$$

From (71) and (82)

$${}_jF_0 = P_{01} V_1 + (1 - P_{01}) V_0. \quad (84)$$

From (73) and (83)

$${}_jF_1 = \rho_{11} V_1 + (1 - \rho_{11}) V_0. \quad (85)$$

Revision of the Nebeker-Moy Model

Now presumably the sum of the coefficients of the V 's in equation (84) should not add up to 1 as they do but, rather, to ρ_0 , the probability that level 0 of the prior dimension will occur. Similarly, it would appear that the coefficients of the V 's in (85) should add up to ρ_1 . These conditions will be satisfied if we write, respectively,

$${}_jF_0 = \rho_{01} V_1 + \rho_{00} V_0 \quad (86)$$

$${}_jF_1 = \rho_{11} V_1 + \rho_{10} V_0 \quad (87)$$

as can be seen by equations (67) and (68), respectively.

However, to make the equations come out as in (86) and (87), we would have to rewrite equations (82) and (83), respectively, as

$${}_jF_0 = \rho_0 P_{01} V_1 + \rho_0 (1 - P_{01}) V_0 \quad (88)$$

$${}_jF_1 = \rho_1 P_{11} V_1 + \rho_1 (1 - P_{11}) V_0. \quad (89)$$

Then using (76) and (77) and going back to equation (65), we would substitute for it:

$${}_jF_i = E_i P_{ij} V_j + E_i (1 - P_{ij}) \bar{V}_j. \quad (90)$$

Finally, instead of equation (64), we would have

$$F_i = \sum_{j=1}^n (E_i P_{ij} V_j + E_i (1 - P_{ij}) \bar{V}_j). \quad (91)$$

But (91) can be written

$$F_i = E_i \sum_{j=1}^n (P_{ij} V_j + (1 - P_{ij}) \bar{V}_j). \quad (92)$$

Possible Sources of Confusion

If it is the case that equation (92) rather than (64) would better suit the logic of Nebeker and Moy (1976), we might inquire into possible sources of confusion that lead one to equation (64) rather than equation (92).

The Algebra of the Model

Let us first go back to Vroom's (1964) original choice model.

$$F_i = \sum_{j=1}^n (E_{ij} V_j) \quad (93)$$

where

- F_i = the force on the individual to perform act i
 E_{ij} = the strength of the expectancy that act i will be followed by outcome j
 V_j = the valence of outcome j, and
 n = the number of outcomes.

Nebeker and Moy recognized the basic weakness of this model in that V_j is meaningless as a dimension unless it is given at least two levels. They assigned the lowest level 0 (nonoccurrence) and the other level 1 (occurrence). Then \bar{V}_j was defined as the valence of the 0 level and V_j as the valence of the 1 level for outcome dimension j. If E_{ij} was the expectancy of act i being followed by level 1 of outcome j, then $1 - E_{ij}$ would be the expectancy of level 0 of outcome j, since the two levels are mutually exclusive but completely inclusive. According to this reformulation, equation (93) would be rewritten

$$F_i = \sum_{j=1}^n (E_{ij} V_j + (1 - E_{ij}) \bar{V}_j). \quad (94)$$

Equation (94) appears to be definitely superior to (64) in logic and rationale but it still presents some rather major philosophical and semantic difficulties.

The left-hand side of the equation, F_i , is defined as the force on the individual to perform act i. Each of these three words is subject to criticism in varying degrees. Perhaps the worst of the three is force. It is not enough to say that its meaning is defined in the right side of equation (94). If this equation aspires to the status of a mathematical model, the least we can do is attempt to operationalize the left-hand side.

Suppose we multiply both sides of equation (94) by E_i , the expectancy that event i will or can occur, whether the event be a level of a dimension or a choice of one among a group of alternatives. We would then have

$$E_i F_i = E_i \left(\sum_{j=1}^n (E_{ij} V_j + (1 - E_{ij}) \bar{V}_j) \right). \quad (95)$$

Aside from the notational substitution of E_{ij} for P_{ij} , the right side of equation (95) is the same as equation (92). The left-hand side of (95) is E_i times the left-hand side of (92). Hence we must assume that F_i cannot have the same meaning in both cases.

Before a mathematical equation can achieve the status of a mathematical model, the variable of the left-hand side must be defined in terms of the operational or experimental procedures required to obtain a value for it. This procedure must be independent of those for obtaining the values in the right-hand side. The criterion variable is on the left-hand side and the predictor variables are on the right-hand side. Obviously, whatever the left-hand side is, it must be something related to level i of the criterion or dependent variable.

Variables that suggest themselves are (1) the value assigned to the level, (2) the expectation that the level will or can occur, (3) the valence of the level, and (4) the proportion of times the event actually occurs at that level.

Each of these four variables can conceivably be evaluated for a given level for some defined dimension of some specified modality. However the variable is defined, it is possible to set up a loss function and apply more or less adequate tests of significance for a single entity or for a modality of entities. However, this observation suggests the issue of within vs. between-group analyses, which we are not yet ready to discuss.

Another unusual characteristic of equations (94) and (95) is that the right-hand side contains no parameters. Presumably, the E_{ij} 's, the V_j 's, and the \bar{V}_j 's are all experimentally obtained. But it is probably not feasible to attempt to introduce parameters into the model until the criterion variable has been adequately operationalized.

In addition to a lack of clarity in the operational definition of the dependent variable, there is another source of confusion in attempting to develop a more meaningful model from Vroom's original choice model. Nebeker and Moy (1976) were entirely justified in recognizing that an outcome, however defined, must have at least two levels. But the concept of multilevel dimensions for both prior and subsequent dimensions introduces a source of confusion with respect to expectancies.

Kinds of Expectancies

We have identified the expectancy that a given level of a prior dimension will occur with a given level of a subsequent dimension. This we call a joint expectancy. We have also identified the expectation that a given level of a prior dimension will or can occur and the expectation that a given level of a subsequent dimension can or will occur. But we can also have another type of expectancies.

Suppose that a given level of a prior dimension has already occurred. Then we may ask, for this level, what the expectancies of occurrence are for each level of the subsequent dimension. Obviously, the sum of these expectancies must be 1. These may be called conditional expectancies.

We see then that, for every pair of prior and subsequent levels, there are two expectancies: the first is the joint expectancy and the second, the conditional expectancy. A conditional expectancy is obtained by dividing the corresponding joint expectancy by the expectancy of the corresponding prior dimension level. It should also be noted that this latter expectancy is the sum of the joint expectancies. These relationships are implicit in equations (67) through (75).

Can and Will Expectancies

A final possible source of confusion may be noted. This concerns the expectancies of occurrence of the levels of the prior dimension. This dimension may be either an effort or a performance dimension, however defined. We may clearly have two kinds of expectancies with respect to the occurrence of a given level: (1) the expectancy that the event will occur, and (2) the expectancy that the event can occur. The expectancy may be very high that the event can occur and yet very low that it will.

Special Nebeker Models

Our discussions of the Nebeker and Moy model, and others as well, indicate the crucial importance of the concept of levels of dimensions in expectancy theory. Further special cases of expectancy theory modeling that emphasize this crucial role have been suggested in personal communication by Delbert Nebeker. Two examples of these are given in Appendices I and J.

THE UNRESTRICTED LEVEL MODEL

Before proceeding further with our analysis of the utility or disutility, as the case may be, of the instrumentality concept, let us consider the general case of the revised choice model. In the original model, the number of levels of the prior dimension was unrestricted. In the Nebeker-Moy model, the number of levels for the subsequent dimension was explicitly increased from the uninterpretable single level to the two-level dimension. We shall now generalize the model so that the number of levels for both the prior and subsequent dimensions are theoretically unrestricted. (Practical considerations, as we shall see, may well impose restrictions on these numbers.)

The Two-dimensional Case

The Mathematical Notation

Suppose we have a dimension of a prior modality with levels l, n_1 and a dimension of a subsequent modality with levels l, n_2 . Let

ρ_{ij} be the probability that level i of the prior dimension and level j of the subsequent dimension will occur,

${}_2V_j$ be the valence of the j th level of the subsequent dimension,

$$\rho = \begin{matrix} \rho_{11} & \dots & \rho_{1n_2} \\ \vdots & & \vdots \\ \rho_{n_1 1} & \dots & \rho_{n_1 n_2} \end{matrix}, \quad (96)$$

$${}_2V = (V_1, \dots, V_{n_2}), \quad (97)$$

$${}_1\rho = \rho \cdot 1, \quad (98)$$

$${}_2\rho = \rho' \cdot 1. \quad (99)$$

Thus ${}_1\rho$ is a column vector of row sums of ρ and ${}_2\rho$ is a column vector of column sums of ρ . We let

$${}_1\rho_i = i \text{ th element of } {}_1\rho$$

$${}_2\rho_j = j \text{ th element of } {}_2\rho$$

$${}_1V = \text{the vector of } n_1 \text{ valences for the } n_1 \text{ levels of the prior dimension } 1.$$

Then ${}_1\rho_i$ is the expectancy that level i of the prior dimension will occur and ${}_2\rho_j$ is the expectancy that level j of the subsequent dimension will occur. According to the definitions

$${}_1\rho' 1 = {}_2\rho' 1 = 1 \quad (100)$$

or the total number of probabilities must add up to 1.

We see therefore that ρ may be thought of as derived from a table of frequencies in a scatter diagram with the numbers of class intervals being n_1 and n_2 respectively. The elements of this table are divided by the total number of observations to give the ρ matrix. Also, the ${}_1\rho$ and ${}_2\rho$ vectors are the marginals of relative frequencies for n_1 and n_2 respectively. We let

${}_1L$ = the n_1 th order vector of the values of the levels for the prior dimension

and

${}_2L$ = the n_2 th order vector of the values of the levels for the subsequent dimension.

We have then the matrix ρ and the vectors ${}_1\rho$, ${}_2\rho$, ${}_1V$, ${}_2V$, ${}_1L$, ${}_2L$. Theoretically, we could also have a matrix of valences V where V_{ij} is the valence of the i th level of the prior dimension when followed by the j th level of the subsequent dimension. This is a conceptually meaningful valence. One could realistically prefer one pair of dimension levels to another. However, for the present, we do not suggest incorporation of valences of such interactions of levels into an experimental design. We shall restrict ourselves to the matrix and vectors just mentioned and see how they might be related to the current instrumentality and expectancy and valence concepts.

The Correlation Between Two Outcomes

First, let us examine the notion of the correlation or instrumentality between two outcomes. We have already considered the case of two levels each for prior and subsequent dimensions and have seen how we may obtain a phi coefficient from the expectancy values generated by them. By way of introduction, we review the procedure by which we may calculate a correlation coefficient for any two variables x and y from grouped data. The ascription of discrete levels to dimensions is precisely what is done when class intervals are set up for experimental variables. We let

- N be the number of cases
- n_x be the number of class intervals for x
- n_y be the number of class intervals for y
- v_x be a vector of elements from 1 to n_x
- v_y be a vector of elements from 1 to n_y
- $m_{x_i y_j}$ be the number of cases in the i th class interval of x and the j th class interval of y

m_{xy} be an $n_x \times n_y$ matrix of the $m_{x_i y_j}$
 l be a vector of 1's whose order is determined by context
 c_{xy} be the covariance between the x and y variables
 σ_x be the standard deviation of the x variable
 σ_y be the standard deviation of the y variable
 r_{xy} be the correlation between x and y.

Then it is well known that

$$c_{xy} = \frac{v_x' m_{xy} v_y}{N} - \frac{v_x' m_{xy} l}{N} \frac{l' m_{xy} v_y}{N} \quad (101)$$

We let

$$f_x = m_{xy} l \quad (101a)$$

$$D_x l = f_x \quad (101b)$$

$$f_y = m_{yx} l \quad (101c)$$

$$D_y l = f_y. \quad (101d)$$

Then it is also well known that

$$\sigma_x = \sqrt{\frac{v_x' D_x v_x}{N} - \left(\frac{v_x' m_{xy} l}{N} \right)^2} \quad (102)$$

$$\sigma_y = \sqrt{\frac{v_y' D_y v_y}{N} - \left(\frac{l' m_{xy} v_y}{N} \right)^2} \quad (103)$$

Finally, of course, we have

$$r_{xy} = \frac{c_{xy}}{\sigma_x \sigma_y}. \quad (104)$$

Expectancies and the Correlation Model

Suppose now we identify analogous symbols from the expectancy formulation just preceding with the correlation formulation just completed. We have

$$n_1 = n_x \quad (105)$$

$$n_2 = n_y \quad (106)$$

$$\rho = m_{xy}/N \quad (107)$$

$$1^{\rho} = \frac{m_{xy} - 1}{N} \quad (108)$$

$$2^{\rho} = \frac{m_{yx} - 1}{N} \quad (109)$$

$$1^L = v_1 \quad (110)$$

$$2^L = v_2 \quad (111)$$

We may also indicate the following translation of notation:

$$c_{12} = c_{xy} \quad (112)$$

$$\sigma_1 = \sigma_x \quad (113)$$

$$\sigma_2 = \sigma_y \quad (114)$$

$$r_{12} = r_{xy} \quad (115)$$

and let

$$D_1 1 = 1^{\rho} \quad (116)$$

$$D_2 1 = 2^{\rho} \quad (117)$$

Then

$$c_{12} = 1^{L'} 2^L - (1^{L'} 1^{\rho}) (2^{\rho'} 2^L) \quad (118)$$

$$\sigma_1 = 1^{L'} D_1 1^L - (1^{L'} 1^{\rho})^2 \quad (119)$$

$$\sigma_2 = 2^{L'} D_2 2^L - (2^{L'} 2^{\rho})^2 \quad (120)$$

and finally, of course,

$$r_{12} = \frac{c_{12}}{\sigma_1 \sigma_2} \quad (121)$$

We see then that we can, by time-honored method, define and calculate a correlation coefficient between dimensions of prior and subsequent modalities, provided that at least two levels are specified for each. We have earlier considered the limiting case of the phi coefficient when only two levels were specified for each dimension. Even for that case, we were appalled by the statistical sophistication and computational wizardry required of the subjects in an actual experimental project. In the case of more than two levels, the task becomes even more difficult. And it is not made any easier by calling the correlation coefficient an instrumentality.

Instrumentality with Unrestricted Levels

There is, however, a more serious criticism of the instrumentality concept than merely the computational burden it places on the subject. We may look again at the valence or instrumentality model as proposed by Vroom (1964). This we repeat for easy reference, as follows:

$$V_j = f \sum_{k=1}^n (V_k I_{jk}) \quad (122)$$

where

V_j = the valence of outcome j

I_{jk} = the cognized instrumentality of outcome j for attainment of outcome k

V_k = valence of outcome k

n = number of outcomes.

Now as soon as we interpret I_{jk} as a correlation coefficient, we imply at least two levels of outcome for both j and k . But now we have both levels of valence and levels of outcome implied for each outcome. To which level of outcome does valence V_k apply? A similar question must be asked about V_j also.

In the case of only two levels for each outcome, one might arbitrarily substitute the valence of level $\underline{1}$ for each of the n outcomes and say that the calculated valence applies only to level $\underline{1}$ of outcome j . But what about the valences of $\underline{0}$ levels of outcome k ? Attempts to give meaning to equation (122) become even more hopeless when one increases the number of levels for outcome j and for the n outcomes k .

There seems, then, little choice but to abandon the concept of instrumentality completely as proposed by Vroom (1964) and with it the model that it purports to implement.

The Multidimensional Case

Let us then return to our revision of the choice model, as suggested in equations (92) or (95), and see how it fares when we generalize the number of levels ascribed to a prior dimension and to the dimensions of a subsequent modality. For the time being, we shall ignore the disembodied spirit on the left-hand side of these equations, called "the force to perform act i ," in the faith that it can be replaced by something more substantial. Instead of considering only one dimension of a subsequent modality, we shall have two or more such dimensions.

Mathematical Development

We let

n_a be the number of levels of a dimension from a prior modality

m be the number of dimensions of a subsequent modality

n_k be the number of levels of the k th dimension of a subsequent modality

k^{ρ}_{ij} be the joint expectancy that level i of a prior dimension and level j of dimension k of a subsequent modality will occur

k^{ρ} be an $n_a \times n_k$ matrix of the joint expectancies k^{ρ}_{ij}

- ${}_k P_{ij}$ be the conditional expectancy that, if level i of the prior dimension occurs, it will be followed by level j of dimension k of the subsequent modality
- ${}_k P$ be the $n_a \times n_k$ matrix of conditional expectancies ${}_k P_{ij}$
- ${}_a \rho_i$ be the expectancy that level i of the prior dimension will occur
- ${}_a \rho$ be the n_a th order vector ${}_a \rho_i$
- ${}_a V_i$ be the valence of level i of the prior dimension
- ${}_a V$ be the n_a th order vector of the ${}_a V_i$
- ${}_k V_j$ be the valence of the j th level of dimension k of the subsequent modality
- ${}_k V$ be the n_k th order vector of the ${}_k V_j$
- ${}_a D$ be an n_a th order diagonal matrix of ${}_a \rho_i$.

Then by definition,

$${}_a \rho = \rho \mathbf{1} \quad (123)$$

$${}_k P = {}_a D^{-1} {}_k \rho \quad (124)$$

$${}_k P \mathbf{1} = \mathbf{1} . \quad (125)$$

Suppose now we define

$${}_k F = {}_a D {}_k P {}_k V . \quad (126)$$

From (124) and (126)

$${}_k F = {}_k \rho {}_k V . \quad (127)$$

Now ${}_k F$ is an n_a th order vector. We then let

F be an $n_a \times m$ order matrix made up of the m dimensions of the subsequent modality

U be an m th order vector whose elements are as yet unspecified.

Finally we let

$$\tilde{Y} = F U . \quad (128)$$

It may not be obvious but, if we let U be a vector of 1's and $n_k = 2$ for all values of k , then \tilde{Y}_i , the i th element of \tilde{Y} , is equivalent to the right-hand side of equation (92).

Practical Considerations

We may now raise a practical issue concerning equations (126) and (127) from which the \tilde{Y} vector in (128) is calculated. Equation (126) involves the matrix of conditional expectancies ${}_k P_{ij}$ and equation (127), the matrix of joint expectancies ${}_k p_{ij}$. The practical issue is whether it is easier for a subject to express his expectancy of a joint occurrence of two levels or his expectancy that a given level of a subsequent dimension will occur, given the occurrence of a specified level of a prior dimension. The answer to this question can probably be resolved experimentally.

The Criterion Modality

But a more fundamental question has to do with the interpretation of \tilde{Y} on the left side of equation (128). If we are to take the basic principles of multivariate mathematical modeling seriously, then \tilde{Y} must be regarded as an estimate of a set of values experimentally determinable for the n_a values corresponding to the n_a levels of the prior dimension. Presumably, the values in F on the right-hand side of (128) are also experimentally determined. The particular problem under investigation must determine the variable which equation (128) is designed to predict.

It should be noted that the general form of the equation and those from which it is constructed are compatible with the objectives of both the valence and choice models of Vroom (1964). If the general expectancy-valence concepts are viable for predicting human behavior, we may be justified in exploring the ramifications of the equation in some detail. Before doing this, however, we shall first take up some of the issues alluded to earlier in this document.

REVIEW OF THE EXPECTANCY THEORY MODEL

We have already discussed the modality elements of the structural components of multivariate analysis systems with special reference to an expectancy theory model. We have tried to interpret the essential modalities of the structure in terms of organism, environment, and interaction of the two. While we have not found the modality elements gleaned from the expectancy theory literature completely adequate from the interaction point of view, we were able, for the most part, to articulate the modalities of the expectancy system with those of the interactive system. To avoid tearing down the expectancy castle so that a completely new interactive structure can be built, we shall stick to the established terminology. This should make for continuity in the expectancy theory literature and avoid the futile practice so common in psychology of discarding current fads and instituting new ones. Along the way, we may issue caveats about unfortunate terminology and reject only the most noxious concepts, such as "instrumentality," whose feet of clay crumble as soon as the light of mathematical analysis is turned upon it.

The Structural Elements

The modalities of expectancy theory we have already discussed are listed below:

Persons

Person characteristics

Evaluators
Organizations
Jobs
Task goals
Contingencies
Effort
Performance
Outcomes
Needs satisfaction
Valence
Expectancy

From this list, the word "motivation" is conspicuous by its absence. This is because, like the weather, everybody talks about it and no one does anything about it. Even the need for the concept does not appear to be clearly established.

Unlike "motivation," personal characteristics, such as aptitude, skill, training, biographical data, and so on, are generally recognized as important for predicting organizational behavior. But, as is the case with motivation, not much of substance has been done to integrate this modality into expectancy theory models. There is some implication of its use when a subject is asked to indicate his expectancy that he would be able to perform at some specified level. Presumably, the notion is that persons of higher ability would expect to be able to perform at higher levels than would persons of lower ability. Self-evaluation of relevant ability to perform seems to be synonymous with expectation.

The issue here seems to be primarily one of the evaluator. The evaluator modality does not appear to have played a major role in expectancy theory models although several studies have used other than self-report evaluations. The most important distinction in evaluator dimensions is probably between that of objective and subjective measures rather than between self reports and subjective reports of others.

The organization modality refers to the particular company or type of occupation chosen by the subject, while the job modality refers more typically to the role selected by or assigned to the person. Each of the modalities can have associated with it a characteristics modality with its dimensions and their levels. Both expectancy and valence levels can be assigned to the levels of the characteristics dimensions for the organization and job modalities. But these two modalities can be of special interest because both valence and expectancy levels can be assigned to them. When this is done, it has been customary to say that the job choice model is being used.

This is an important issue because dimensions of the organization and job modalities may be used directly in equation (89) without reference to level. The distinction is between modalities whose dimensions do not have levels and those that do. Usually a modality whose dimensions do not have levels associated with it can have associated with it a characteristics modality whose dimensions do have two or more levels. For example, the dimensions of a job modality may be all the jobs designated within a given company. The dimensions of a characteristics modality may consist of a taxonomy of aptitudes,

some of which are required in varying degrees (levels) for successful performance in the different jobs.

The task-goal modality has played a major role in expectancy theory models but it has not been emphasized as a focal part of the working environment and is difficult to concretize without the concept of a pre-operant and post-operant structure. It has not been well identified as a separate component in any mathematical model of expectancy theory.

The contingency modality has been used by Nebeker and Moy (1976) with dimensions that could have two or more levels each, but it has not been incorporated into any overall mathematical model. How it can be articulated with the valence and expectancy concepts still remains to be seen. It is an important modality for organizational behavior and merits more than incidental treatment in a comprehensive model. Unlike the person characteristics modality, it is probable that it can be incorporated in an expanded model without recourse to other than the self-report dimension of the evaluator modality.

The effort modality is among the most utilized of the expectancy model modalities. It has come under attack because of problems of operationalization, but it appears to be a crucial link in the chain of expectancy theory modalities. In a sense, it is the only modality identifiable directly with the operant modality of the organism. As such, it should be strengthened rather than relegated to a minor role in an overall model.

The Performance Modality

Perhaps the most intensively exploited modality in the expectancy model chain is the performance modality. This is not an entirely fortunate situation in view of the somewhat ambiguous status of the concept. "Performance" would seem to imply some sort of operation or activity but, if we agree to identify effort with the environment-modifying role of the organism, we must look elsewhere for whatever it is that performance can be identified with. There seems to be little left but the focal environment which the organism by its effort has modified into a restructured focal environment. Presumably, what is really meant by performance is not an act or activity but an accomplished "task goal" or product. It would probably be better to replace the word "performance" with product or production.

It is true that we speak of an actor or entertainer as giving a good "performance." And that is precisely what we should do, for it is the activity--actual physical movements--to which we are referring. But, in the organizational setting, we are not concerned with the verve and grace, the theatrics of the worker in creating the product. Rather, it is the quality of the product, or the dispatch with which it was completed, that is of concern to the organization objectives. The entertainment value of the worker while producing the product is at best of only secondary interest in most industrial settings.

The Outcome Modality

Perhaps the most treacherous of all modality designations found in expectancy theory literature is that of "outcome." It means so much that it means very little. It is not enough to say that an outcome is anything an individual might want to attain. We may hazard a more restrictive and perhaps more useful definition from the point of view of human behavior modification. An outcome is a valent restructuring of a person's relevant environment, following either his own operant activity or the restructured focal environment resulting from it. The restructured environment may include dimensions from his external, internal, or ambient environment.

The Needs Modality

As we noted earlier, the needs satisfaction modality is rarely utilized in expectancy theory modeling although it might eventually prove useful. It is sometimes referred to as a second-level outcome. This is a poor use of the term unless we can find a better one to use in referring to the levels of a dimension. Even at that, "second level" implies that it is above or below in magnitude when this is, presumably, not at all what is meant. Probably needs satisfaction follows valent environmental restructuring (job outcomes) so that all we really require is an indication of modality ordering or sequence. And this requirement is applicable to all the links in the chain of expectancy theory modalities and not just to the needs satisfaction modality. For this reason, we have suggested the terms prior and subsequent modalities.

The Valence Modality

We have already discussed the valence modality at some length and have suggested that it might be multidimensional rather than unidimensional. We shall see later that the transitivity issue might be resolved by such an expansion. We raise again the proposal of some that valence might be defined as anticipated value. We suggested that, until the distinction between current and anticipated value might be found useful, it should not be made. However, the distinction does have some interesting possibilities. In the first place, anticipation does have an implication of expectancy. The concept of the expectancy of the valence of an event might prove useful in the formulation of a more comprehensive expectancy theory model.

But the distinction also may resolve a logical difficulty that current models encounter. A segment of the chronological chain of expectancy modalities proceeds from performance to valent outcomes. We wish to predict performance level from valent outcomes. But we cannot predict an event from another that has not yet occurred. Suppose, however, we conceptualize the valent outcome as the perception of the value of the outcome not only prior to its occurrence but also prior to the occurrence of a level of performance. Thus, we reverse the order of the predictor variable (outcome) and the criterion variable (performance), so that the former precedes the latter in time, as it should.

The Expectancy Modality

The expectancy modality is probably the most fundamental and is the one from which the model takes its name. Aside from the concepts of risk and certainty, there has been little speculation as to the need for more than one dimension. There is, however, the most important distinction between the expectation that an event can occur and that it will occur. A person might have a high expectancy that he can get a certain job if he wants it but a low expectation that he will get it for the simple reason that he does not want it. He may have a high expectancy that he can produce at a high level but a low expectancy that he will because he is convinced that it is not worth while. It is important to recognize the difference between these two kinds of expectancies and the implications of each in expectancy theory model building.

DIMENSIONS OF MODALITIES

Having reviewed the major modalities gleaned from expectancy theory literature, we are now ready to discuss in more detail the subject of dimensions and modalities. We have already called attention to some modalities whose dimensions do not have levels, and

some for which the need of more than one dimension has not been clearly established. But, for some of the modalities, it is relatively easy to suggest a large number of dimensions.

Levels of Dimensions

Perhaps a helpful criterion for identifying modalities whose dimensions do not require levels is that of homogeneity. Certainly the person modality consists of relatively homogeneous dimensions; namely, persons. Assuredly, persons may vary with respect to many different attributes or characteristics. But, having said that, we have introduced another modality, viz., a person characteristics modality. And this modality may include a vast array of highly heterogeneous dimensions, each of which do have two or more levels.

Similarly, the evaluator modality may have a number of different dimensions but be homogeneous with respect to their function; namely, that of assigning levels of other dimensions.

The organization and job modalities have dimensions without levels, but corresponding to each may be a characteristics modality whose dimensions do have levels. Levels of the valence and expectancy modalities may be assigned both to each of the dimensions of these modalities and to the levels of the dimensions of their characteristics modalities. A person may indicate a strength of preference (valence) for a company or a kind of job or his liking (valence) for various degrees of specified characteristics of a company or job. Similarly, he may indicate his expectancy that he could or would work in a given company or at a given job. He can also indicate his expectancy that, for a given dimension of a characteristics modality for either modality, a given level of that dimension would obtain.

Homogeneity of Dimensions

In general, we may examine a modality to see whether its dimensions are homogeneous or whether it is a characteristics modality of heterogeneous dimensions associated with the modality of homogeneous dimensions. If it is a modality of homogeneous dimensions, the dimensions will not in general have levels, but levels of both valence and expectancy may be assigned to a dimension for a person by an evaluator. If it is one of the heterogeneous dimensions of a characteristics modality, the dimension has at least two levels, and both a valence and an expectancy level may be assigned by an evaluator to each level for a person.

This kind of modality analysis has not been considered in the expectancy theory literature even by implication. It is probable that a more adequate expectancy theory model could benefit substantially from a systematic analysis of modality types.

Our chief concern at this point is to identify heterogeneous dimension modalities that are of great practical importance for expectancy theory models. Considerable attention has been given to this problem in the literature, particularly with reference to the dimensions of organization and job characteristics modalities, contingency modality dimensions, task goal dimensions, and outcome and needs satisfaction dimensions. The concern over the specification of these dimensions is usually expressed as the need for adequate taxonomies.

Specification of Dimensions

The practical approach to dimension specification is to draw up check-lists of things that people like or dislike about their jobs or of things they want as a result of working at the job. Such lists are generated by interviews with workers, administrators, and so on, and by investigators themselves. A number of practical as well as theoretical considerations are involved in the generation of such check lists.

The Comprehensiveness of Dimensions

Perhaps the most important of these considerations is how adequately and completely a list covers the important factors for all persons. For the contingency modality, for example, does the list cover most of the things about the ambient environment that help or hinder the person from doing his job (adequately restructuring the focal environment)? The same question can be asked about the list that describes those things about the ambient environment, the company or job, that make it a more or less desirable place in which to work but that do not bear directly on job performance.

Perhaps the list of job outcomes has been given more attention than that of any other link in the chain of modalities. This is, of course, because this modality plays such a central role in current expectancy theory models. The items in this list represent the variables whose valences and expectancies are used to predict performance, or "force to act," or job choice. This list is the foundation for both the valence and choice models proposed by Vroom.

A major problem in generating a list of outcomes or dimensions of other modalities is that, if the list is extensive enough to assure adequate coverage, it is so long that the task of responding to all the items becomes too great. Another objection sometimes raised to a comprehensive list of outcomes or dimensions of other modalities is that only a relatively small number of the items may be relevant for most persons in job situations.

This latter objection has been solved in part by an ingenious technique by Nebeker and Moy (1976). The subject is first asked to go through a previously compiled comprehensive list and indicate the five items which he considers the most important outcomes of job performance or effort. He also ranks these five in order of importance. In further responses involving the outcomes, only his own five are considered by any one person, thus greatly simplifying the task of the subject.

Several assumptions appear to be involved in this technique. The first has to do with the concept of valence, which may be either positive or negative. In fact, "importance" appears to be used synonymously with "valence." There has been some discussion in the literature on whether "importance" and value should be considered as two separate dimensions. One conclusion is that they should be treated as separate until proven experimentally to be the same. This conclusion can be extended to absurd lengths. To insist that any two words must be assumed to have different meanings until proved to be the same imposes a ridiculously impossible, not to say unnecessary, task on all investigators whose subject matter depends heavily on language or communication symbols. All scientists, particularly social scientists, are constantly under the necessity of yielding to realistic considerations and making practical compromises. It can, in fact, be documented that it is more often the case that things that seem different are essentially the same than that things that seem the same are essentially different. The empirical evidence for this is in studies that have attempted to identify new aptitude or personality dimensions. Usually the hypothesized new dimensions tend to collapse into previously established ones, regardless of whatever respectable statistical analysis that is used.

A second assumption implied by the Nebeker technique is that the items that were not checked as among the most important may be regarded as having zero valence for the individual. Therefore, no matter what expectancy level the valence of the item was multiplied by, it would vanish in any model based on the products of valences times expectancies. An assumption subsidiary to this emerges from the explicit introduction by Nebeker of a two-level outcome (present or absent). Even if the number of specified levels is increased beyond two, the subsidiary assumption is that, for items not in the "important" group, all levels may be treated as having zero valence. The assumption of zero valence for items not included in the "important" group appears justifiable, particularly if the number in the "important" group is sufficiently large.

The Specificity of Dimensions

Another issue frequently considered in the generation of a list of outcomes or dimensions of other modalities is the degree of specificity of the items. There are two questions related to this issue:

1. What should be the general level of specificity of all the items admitted to the list?
2. What range or dispersion of specificities should be tolerated?

These are practical questions that cannot be answered in the abstract. But perhaps there has been more agonizing over these questions than the issue merits. These are not issues peculiar to expectancy theory models. They permeate all areas of psychological measurement technology, particularly those concerned with the vast array of noncognitive inventories, both commercial and amateur, that have flooded the social science arenas for more than half a century.

Dimensions and Verbalizations

At this point, we may offer a defense for checklists, rating scales, inventories, and all manner of aggregations of verbal stimulus elements designed to elicit quantifiable responses from persons about themselves or other persons or things. There is a tendency for investigators to be apologetic about such approaches to psychological measurement. We suggest that there is no justification for this implied distrust of reliance on verbal communication symbols as the materials from which to construct scientifically valid quantification instruments. We suggest that a gingerly, diffident, and overcautious approach to the generation of verbal stimulus elements for psychometric purposes is easily and often overdone. The collection of such elements should more often than not consist of spontaneous unedited verbal responses of real live people in real world situations and activities.

Items that are generated in this way and from the sources suggested almost certainly will exhibit a wide range--from specificity to generality. They may not be elegantly phrased but, in general, they may be expected to be more meaningful to the layman or person in work situations than the most polished or erudite phrases or sentences that the investigator can fashion.

But, to carry the matter to a more general level, reluctance to exploit verbal stimulus patterns or communication media to the fullest in behavioral sciences programs seems unjustifiable for a most basic reason. Perhaps the most important difference between man and other mammals is the far superior communication facilities of man and the implications of these for the control of and adjustment to his environment. This

facility encompasses and reflects a very large part of man's environment. This facility encompasses and reflects a very large part of man's socially and personally meaningful activity. When viewed in this light, the criticisms of and apologies for the extensive use of verbal materials for human behavioral research in general and for the fabrication of quantification instruments in particular lose their force. Even the seismologists have at best their Richter scale and the minerologists must make do with their hardness scale.

Latent vs. Manifest Dimensions

We have gone to considerable length to defend the manner in which verbal lists are generated to represent the dimensions of the modalities in expectancy theory. We have also suggested that there may be undue concern about the issues of coverage and specificity. But we must recognize that, having generated a list of dimension statements, there are genuine issues to be resolved.

One of these has to do with redundancy, or the extent to which items in a list of dimensions may overlap or duplicate one another. A list that is elicited from persons in work situations is bound to contain many redundant statements that expand the length of the list without contributing additional information. This leads us into the subjects of taxonomy and of manifest vs. latent dimensions.

When investigators talk about the need for taxonomies, what they imply, even though they may not recognize it, is the need for a list of dimensions of a modality that are in some sense basic or fundamental. A list that is generated by survey or interview methods may be called the manifest dimensions of the modality. To be sure, the list may undergo some editing and a weeding out of obvious duplications and highly specific or trivial items.

Factor Analytic Methods

Even such a list must still be regarded as the manifest dimensions of a modality. The more sophisticated proposals for arriving at a more basic or fundamental set of dimensions have recommended factor analytic techniques. Such techniques may be regarded as yielding the latent dimensions that were discussed earlier in this document. They also may provide the basis for a logical and parsimonious taxonomy for the dimensions of a modality. In general, these techniques can be expected to yield a number of latent dimensions much smaller than the number of manifest dimensions from which they were derived, without losing any of the essential information provided by the original list of dimensions. Also, if competently done, they will yield relatively independent dimensions.

But a number of issues must be considered in the factor analytic approach to the generation of an acceptable taxonomy. One cannot, as is often implied, merely wave the magic wand of factor analysis to solve otherwise insoluble problems for the simple reason that there is no magic wand of factor analysis. There are many different kinds of factor analysis and one cannot merely consult an index of BiMed computer programs with a reasonable expectation of finding the right one.

There are two basic parts to a factor analytic solution. One consists of determining the latent dimensionality of the manifest system and the other is to find a unique coordinate system for the latent variables. The first is the problem of finding an arbitrary factor loading matrix and the other is the problem of rotation to simple structure. For each of these, a confusing number of techniques is available and a suitable choice implies a higher-than-average degree of statistical sophistication. Nevertheless,

adequate procedures and advisory sources are available and the use of proper techniques presents no insurmountable obstacles.

A more fundamental question about the factor analytic approach to logical dimension reduction concerns just where in the expectancy model program the technique should be introduced. In the notation P_{ij} , V_j , which plays a central role in expectancy theory models, the V_j is the valence level of a modality dimension and the P_{ij} is an expectation that this level will follow the i th level of a prior dimension. These values presumably are assigned during the same experimental session by the subject. The V_j must refer to either a manifest or a latent dimension and the P_{ij} must refer to the same dimension as the V_j .

Limitations of Factor Analysis

We may seriously question the validity of using manifest dimensions in expectancy theory models, even though the invalidity of this procedure might invalidate to a greater or lesser degree most published expectancy theory research. If, therefore, we insist that the product terms should apply to the latent dimensions, we imply that these dimensions must be established prior to and independently of the application of the expectancy model to an experimental setting.

This calls for a separate study to establish the latent dimensions from the manifest dimensions. It means that we must find the correlations among the manifest dimensions in a previously established list. One could find the correlations among the dimensions of a modality of homogeneous dimensions such as organizations or jobs, each of whose dimensions, as such, may be assigned valence levels by each member of a group of persons. This procedure would yield a data matrix whose rows are persons and whose columns may be jobs. An entry of the table would be the valence of a given job for a given person. Such a matrix provides the basis for a factor analysis from which one might select a limited number of dimensions from the original jobs to represent the whole group of jobs.

This procedure does not help us to reduce the dimensions of the outcome modality, as we shall see. Furthermore, it can do as much or more harm than good for the job choice model. In this model, investigators have preferred to increase rather than decrease the number of choice alternatives. This preference is due to a rather well established interest in the general problem of across-subjects vs. within-subjects analyses. The thought seems to be that the more choice alternatives available, the more stable or meaningful or whatnot will be the within-subjects analyses. Whatever the merits of this desire for more choice alternatives, it cannot be satisfied by factor analytic techniques applied to the modality of dependent variables.

Turning to the outcome modality, the application of factor analytic techniques presents different problems. We could, of course, follow Vroom and treat outcomes as dimensions that do not have levels. A list of manifest outcomes could be presented to a group of persons, each of whom would record a valence level for each manifest dimension. Such a procedure would yield a data matrix to which an appropriate factor analytic technique could be applied. This approach is the same as the one we discussed for the organization and job modalities. However, it ignores the concept of even a two-level dimension (occurrence or nonoccurrence) that we have recognized as a step forward in expectancy theory models.

This discussion may lead us to question whether the valence variable is the appropriate one on which to base a factor analysis of manifest outcome dimensions. Actually, what we are primarily concerned with is the reduction in overlapping or redundancy in the list of outcomes. But this concern implies correlations among outcomes, and correlations among dimensions of any modality imply two or more levels of the dimensions. We could formally indicate the meaning of the correlation between two outcomes as we did in our discussion of the concept of the instrumentality of one outcome for another, but it will be recalled that this interpretation implies an expectation level of the joint occurrence of each level of one outcome with every level of the other outcome. If each outcome had only two levels (0, 1), the assignment of tetrads of such expectations for all possible pairs of outcomes might not be prohibitively onerous for a person, provided the list of outcomes was not too long. One could then calculate phi coefficients or tetrachoric correlations for each pair of outcomes and thus get a matrix of correlations among the outcomes corresponding to each person.

Aside from a simple-minded averaging of these matrices for all persons, what rationale can we come up with for consolidating them into a single matrix? Calculation of such matrices is not tenable for two reasons. First, even though we had only ten outcomes with 2 levels each, this would call for 45 correlation coefficients, each based on 4 expectancy estimates, or a total of 450 expectancy estimates for each person. Though perhaps not prohibitive, it might still be highly impractical. In the second place, no elegant rationale for consolidating the estimates for all persons has been proposed.

Whether or not it is feasible, necessary, or even desirable that a taxonomy of outcomes be established by factor analytic or other methods is not yet clear. Before concerning ourselves at length with this issue, we may consider possibilities for circumventing it.

Dimensions and the Needs Modality

First we will comment briefly on the problem of a needs-satisfaction taxonomy. Perhaps the most ambitious effort to develop the dimensions of a basic-needs modality is that of Henry Murray (1938). To date, there has been no effort to integrate the needs-satisfaction modality into a complete mathematical model of expectancy theory. So far, of the multidimensional modalities, only the outcomes modality has been incorporated into a mathematical formulation of the expectancy theory model. As noted earlier, while data have been collected for the person characteristic, job characteristic, and contingency modalities, none of these has been incorporated in a multivariate mathematical model.

Whether any or all of these modalities would have more promise for the development of a comprehensive model than the needs-satisfaction modality cannot even be reliably conjectured at this time, but this latter seems to fall more easily into an expanded model than the others. There is probably no need from a mathematical point of view to distinguish the needs modality from the outcome modality. The dimensions of the two could be combined and all included as valence-expectancy product terms in the mathematical equation. The advantage of considering the dimensions of the needs modality separately might be that the list is probably more general, and might be included in a wider variety of organization or job settings, than outcome lists that are generated for specific research projects.

A great deal of work has been done with Murray's postulated needs. In addition to the work of Murray (1938) and his students, Edwards (1953) has developed a 12-dimensional Personal Preference Schedule, based on Murray's postulated needs, that has provided the

basis of further research. Also Campbell (1959) has developed items purporting to measure Murray's postulated need dimensions and has administered them to groups of subjects. Factor analytic techniques have been applied to the data and the results indicate considerable overlapping of the postulated needs.

Summary

This discussion of the dimensions of modalities has emphasized the distinction between manifest and latent dimensions and the felt need for more adequate taxonomies in expectancy theory modeling. The role of factor analysis in the establishment of objective and parsimonious taxonomic structures has been indicated. We have not attempted to resolve the issues raised in this discussion, but it is important to recognize that the taxonomic problems encountered in expectancy theory formulations are not peculiar to expectancy theory alone. Rather, they are just a small part of the much larger problems that continue to bedevil psychology in all of its branches. In fact, taxonomic problems are common to all scientific disciplines, even though they may be more acute in the social sciences.

REVISION OF VROOM'S MODEL

We shall now return to consider in more detail how the taxonomic issue might be circumvented in the revised mathematical model based on Vroom's original equations.

The Mathematical Development

The development that follows does not necessarily attempt to distinguish between outcomes and needs, and assumes manifest outcomes.

Definition of Notation

Suppose we let k^d be defined by

$$k^d \cdot 1 = k^V. \quad (129)$$

From (127) and (129) we then can write

$$k^F = k^{\rho} k^d \cdot 1. \quad (130)$$

Suppose now we write

$$k^R = k^{\rho} k^d. \quad (131)$$

We note now that the ij th element of k^R is given by

$$k^{R_{ij}} = k^{\rho_{ij}} k^{d_j} \quad (132)$$

Equation (132) means that the expectancy that the i th level of the prior dimension will be followed by the j th level of the k th subsequent dimension is multiplied by the valence of the j th level of the k th subsequent dimension to give the ij th element of k^R .

The Mathematical Formulation

From (130) and (131) we can write

$${}_k^F = {}_k^R \mathbf{1}. \quad (133)$$

Now, instead of (133), we may write

$${}_k^F = {}_k^R {}_k^W \quad (134)$$

where ${}_k^W$ is a vector of weights to be determined. We may regard ${}_k^{R_{ij}}$ as an expected valence since it is an expectancy multiplied by a valence. Therefore, we may regard ${}_k^R$ as a matrix of expected valences. Equally well, we might call it a matrix of valent weighted expectancies.

Now if we go back to equation (128), we may rewrite it, because of (127), as

$${}_a^F = ({}_1^{\rho}, \dots, {}_k^{\rho}, \dots, {}_m^{\rho}) \begin{bmatrix} {}_1^V u_1 \\ \vdots \\ {}_k^V u_k \\ \vdots \\ {}_m^V u_m \end{bmatrix} \quad (135)$$

where, on the right-hand side, the ${}_k^V$ are vectors and the u_k are scalars. If we use (133) instead of (127), we can write (135) as

$${}_a^F = ({}_1^R, \dots, {}_k^R, \dots, {}_m^R) \begin{bmatrix} {}_1 u_1 \\ \vdots \\ {}_k u_k \\ \vdots \\ {}_m u_m \end{bmatrix} \quad (136)$$

where the ${}_k^R$'s are submatrices of expected valences.

Introduction of the Parameters

Now instead of the unit summing vector on the right of (133) and those on the right of (136), let us substitute the vectors of unspecified weights indicated by ${}_k^W$ in (134). Then we get from (136)

$$\tilde{Y} = ({}_1^R, \dots, {}_k^R, \dots, {}_m^R) \begin{bmatrix} {}_1^W u_1 \\ \vdots \\ {}_k^W u_k \\ \vdots \\ {}_m^W u_m \end{bmatrix}. \quad (137)$$

Next let

$${}_k^W = {}_k^W u_k. \quad (138)$$

From (137) and (138), we have

$$\tilde{Y} = ({}_1R, \dots, {}_kR, \dots, {}_mR) \begin{bmatrix} 1^w \\ \vdots \\ k^w \\ \vdots \\ m^w \end{bmatrix}. \quad (139)$$

Finally, suppose we define

$$R = ({}_1R, \dots, {}_kR, \dots, {}_mR) \quad (140)$$

$$w' = (1^w, \dots, k^w, \dots, m^w). \quad (141)$$

From (139), (140), and (141)

$$\tilde{Y} = R w'. \quad (142)$$

Now R in (142) may be called a row supermatrix of joint expected valences because of the definition of the p 's. If we wish to write equation (142) based on the conditional expectancies ${}_kP_{ij}$, we may write

$$R = {}_aD P d \quad (143)$$

where

P is the m th order row supermatrix of the conditional expectancies

and

d is the m th order super diagonal matrix of valences.

We let

$$G = P d. \quad (144)$$

Then G may be called a row supermatrix of conditional expected valences. We may now write

$$\tilde{Y} = {}_aD G w'. \quad (145)$$

Equation (145) is to be compared with (142). It will be recalled that the question of whether conditional or joint expectancies are psychologically easier to estimate is as yet unresolved experimentally, and there seems to be no well established theory to suggest an answer.

Extension to the Person Modality

In any case, it must be recalled that the equations we have been working with are based on only a single person. We must now extend the analysis to all n_a subjects. It will be notationally simpler to work with equation (142) than with (145).

The Matrix Formulation

We shall rewrite equation (140) with additional prescripts to indicate that it refers to the h th person, thus:

$${}_h R \equiv ({}_{h1} R, \dots, {}_{hk} R, \dots, {}_{hm} R). \quad (146)$$

We let n be the number of persons. We rewrite (142) as

$${}_h \tilde{Y} = {}_h R w \quad (147)$$

to indicate that it refers to the h th person.

Now we let

$$R \equiv \begin{bmatrix} {}_1 R \\ \vdots \\ {}_h R \\ \vdots \\ {}_n R \end{bmatrix}. \quad (148)$$

and

$$\tilde{Y} \equiv \begin{bmatrix} \tilde{Y}_1 \\ \vdots \\ \tilde{Y}_h \\ \vdots \\ \tilde{Y}_n \end{bmatrix}. \quad (149)$$

Then the general equation is

$$\tilde{Y} = R w. \quad (150)$$

To expand the notation, we have

$$\begin{bmatrix} \tilde{Y}_1 \\ \vdots \\ \tilde{Y}_h \\ \vdots \\ \tilde{Y}_n \end{bmatrix} = \begin{bmatrix} {}_{11} R, \dots, {}_{1k} R, \dots, {}_{1m} R \\ {}_{h1} R, \dots, {}_{hk} R, \dots, {}_{hm} R \\ {}_{n1} R, \dots, {}_{nk} R, \dots, {}_{nm} R \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_k \\ \vdots \\ w_m \end{bmatrix}. \quad (151)$$

The order of each ${}_h \tilde{Y}$ vector is n_a , the number of levels of the dimension from a prior modality.

The order of each ${}_{hk} R$ is $n_a \times n_k$ where n_k is the number of levels of the k th subsequent dimension.

The order of ${}_k w$ is n_k .

There are n ${}_h \tilde{Y}$ vectors, one for each person.

There are m ${}_k w$ vectors, one for each dimension of the subsequent modality.

The simple order of \tilde{Y} in equation (150) is $n_a \times n$.

The simple row order of R in (150) is also $n \times n_a$.

The simple column order of R in (150) is $\sum_{k=1}^m n_k$.

The simple column order of w in (150) is also $\sum_{k=1}^m n_k$.

Let us assume now that ${}_h Y_i$ is the proportion of time that person h actually performs at level i . Let us assume that ${}_h \tilde{Y}_i$ is the estimate of the proportion of time that person h performs at level i . The ${}_{hk} R_{ij}$ is the joint expected valence for person h of level i of the prior dimension when followed by level j of subsequent dimension k . The ${}_h Y_i$ and the ${}_{hk} R_{ij}$ are presumed known experimentally and we wish to determine the w vector so that \tilde{Y} will in some specified sense best estimate Y .

The Least Square Solution

We can write then

$$\epsilon = Y - R w \quad (152)$$

where ϵ is a vector of residuals and w a vector of parameters to be determined. This is the matrix form of the equation for the multiple regression model that we discussed earlier in our development of the multivariate system. The least square solution for w in (152) is well known to be

$$w = (R' R)^{-1} R' Y. \quad (153)$$

Actually, it may be recognized that, traditionally, a constant vector will be included in the residual equation so that, instead of (152), we would have

$$\epsilon = Y - (I, R) \begin{bmatrix} A \\ w \end{bmatrix} \quad (154)$$

where A is a scalar parameter.

Implications of the Generalized Solution

Now equation (152) is apparently new to expectancy theory models, although in the literature vague references are made from time to time about the use of multiple regression methods. We may discuss some of the implications of this equation.

1. The parameters of the solution. In the first place, equation (152) involves the parameters A and w . As we indicated earlier, multivariate systems typically included parameters to be solved for, although these have been conspicuously absent from current expectancy models.

To understand the meaning of these parameters, we go back to equation (133). Each element of ${}_k F_i$ is the sum of products of expectancies by valences for the i th level of the prior dimension. The summation is across levels of a subsequent dimension. In the simple Vroom (1964) model, there is only one term since his choice model does not explicitly assume levels of dimensions. In the Nebeker-Moy (1976) case, the summation is over two levels (0 and 1). However, since they use conditional expectancies that total to 1 for any level of the prior modality, one of the two expectancies is P_{ij} and the other is $1 - P_{ij}$, and the valences of the 1 and 0 levels are, in their notation, V_j and \bar{V}_j , respectively.

2. The level parameters. If we proceed to equation (134), we see that the summing vector 1 is replaced by a vector of weights (parameters). It is important to note that these parameters are assumed to reflect some characteristics of the levels of the subsequent dimension and are the same for all persons, while the elements in ${}_k R$ are characteristic of the person and vary from one person to another. They are functions of his own expectancies and valences.

One rationale for allowing differential weights in the summation over levels is that the established or specified levels may be arbitrary and not comparable among themselves. Thus, there should be opportunity in the model for a determination of differential values in accordance with their predictive values. The levels may correspond roughly to what we have discussed earlier as ordinal scales where the intervals of a scale are not necessarily equal. We shall have more to say on this subject when we discuss the mechanics of establishing levels for dimensions.

3. The dimension parameters. So far we have considered only the weights or parameters for the levels of a subsequent dimension. We also must combine the results from the separate subsequent dimensions. Here again, instead of a simple sum, we assume a weighted sum, as indicated by equation (137). Here the ${}_k W$ on the right-hand side are vectors of weights applied to levels within a dimension, while the u_k are scalar quantities applied to the dimensions within a subsequent modality.

These u_k weights, as well as the ${}_k W$ weights, are usually taken as unity in the current expectancy theory models. However, for predicting a criterion, there is no reason to assume that all the dimensions of an outcome modality would be of equal weight. The flexibility introduced by the u_k weights effectively dispels much of the anxiety expressed about the comparability and specificity of items on the outcome list. Such concerns are at best no more than academic if the mathematical model includes parameters for weighting any given set optimally for a specified criterion.

4. Combining Levels and Dimensions. We see next from equations (138) and (139) that, mathematically, there is no need for keeping the W vectors and u scalars as distinct values. The two are merged in the k_w vectors of (139) or simply as w in the compressed notation of equation (142). This equation shows that each value of the estimate \hat{Y} of the criterion is a weighted sum of the conditional expected valences of all levels of all dimensions of the subsequent modality.

A Concrete Example

We may now consider a concrete example of the model we have just discussed to get a feeling of its computational and statistical implications.

1. The experimental project. We shall take the data from a study by Nebeker on 128 subjects. This was a job-choice experiment that included 7 different jobs. The outcome modality included 24 different outcome statements. Each outcome had two levels (1 or 0). To simplify both the task of the subjects and the analysis of the data, the subjects were instructed to rate the five most important outcomes, A, B, C, D, and E, respectively. It was assumed that the remaining 19 could be treated as having zero valence and hence could be neglected. Several questions might be raised and suggestions made with reference to this procedure and the data analyses but these matters are not relevant for this discussion. We shall consider only how the model we have been discussing would apply to the data.

2. Degrees of freedom. We have two levels for each of 24 subsequent or outcome dimensions, which gives 48 separate weighting parameters to be solved for. We have 7 jobs and 128 persons, which gives a total of 896 observation equations. A criterion value for a typical observation equation is the expressed valence of a given subject for a given job. A predictor variable consists of a given level of a given outcome. The value of this variable for the criterion value just identified is the expectancy of the given person that the given job would lead to the given level of the given outcome, multiplied by the valence of this level of the outcome for the given person.

If we include an additive constant as indicated in equation (154), we have 49 unknowns to be solved for from 896 observations. Actually, even with the large number of parameters, we have many degrees of freedom and we might expect a reasonably high degree of stability in a conventional least square solution for the 49 parameters. It has not been common to have so many independent variables in a multiple regression analysis but, with generally available computer facilities, the computational burden would be light.

3. Alternative solutions. However, for much smaller numbers of cases and fewer job choices, the number of observations could be much less. For example, suppose we had only 4 job choices and 15 subjects. This would provide only 60 observation equations against 49 parameters to be solved for. One can always wage the crusade for more and more cases but to restrict research to only those areas in which large numbers are available can greatly limit the amount of needed research accomplished. Therefore, other promising techniques for increasing the degrees of freedom of a solution might be considered. One of these is the reduced rank regression analysis approach (see Leiman, 1951; Burket, 1964; Horst, 1970).

Expectancy Theory and the Multivariate Analysis Model

In any case, we have seen how an expectancy theory model may be modified to bring it into the mainstream of multivariate analysis systems. The example just discussed involved the job choice model but it is equally applicable to the performance level

paradigm where the criterion is based on the proportion of the time a person works at a specified performance level. The example includes the structural and functional components of a multivariate analysis system, although only a few of the structural elements of the expectancy model we have discussed at length earlier are present. It includes as alternatives the single dimension performance modality with a number of levels or the job modality with a number of single-level dimensions. It also includes an outcome modality, each of whose dimensions has two or more levels.

In the functional component, we have the equation of the mathematical model, the observation equation, the loss function, the parameters, and the solution for the parameters. This component does not include constraints on the parameters. However, since the model turns out to be a conventional multiple regression model, significance tests are readily available under more or less valid assumptions.

The Within-subject vs. Across-subject Analysis

There is a distinctive feature of the model we have discussed that may be introduced at this time because of the major role it plays in current expectancy theory models.

The Vector-to-Matrix Transition

Suppose we return to equation (149). We let n_i be either the number of levels of a unidimensional criterion modality or the number of dimensions in a criterion modality whose dimensions have only one level. Then the typical vector element ${}_h\tilde{Y}$ can be expanded to simple form, thus:

$${}_h\tilde{Y} = \begin{bmatrix} \tilde{Y}_1 \\ \vdots \\ \tilde{Y}_{n_i} \end{bmatrix} \quad (155)$$

We now arrange the subvectors of (149) in matrix form to be the column vectors of a matrix \tilde{y}' , thus:

$$\tilde{y}' = ({}_1\tilde{Y}, \dots, {}_m\tilde{Y}) \quad (156)$$

In expanded simple form, (156) would be

$$\tilde{y}' = \begin{bmatrix} {}_1\tilde{Y}_1 & \cdots & {}_n\tilde{Y}_1 \\ \vdots & \ddots & \vdots \\ {}_1\tilde{Y}_{n_i} & \cdots & {}_n\tilde{Y}_{n_i} \end{bmatrix} \quad (157)$$

For convenience in further reference we shall reorient (157) to natural order so that

$$\tilde{y} = \begin{bmatrix} {}_1\tilde{Y}_1 & \cdots & {}_1\tilde{Y}_{n_i} \\ \vdots & \ddots & \vdots \\ {}_n\tilde{Y}_1 & \cdots & {}_n\tilde{Y}_{n_i} \end{bmatrix} \quad (158)$$

By analogy we can write a matrix of actual criterion values as

$$y = \begin{bmatrix} 1Y_1 & \cdots & 1Y_{n_i} \\ \cdots & \cdots & \cdots \\ nY_1 & \cdots & nY_{n_i} \end{bmatrix} \quad (159)$$

Now the rows of (158) and (159) represent persons and the columns, choice alternatives such as performance levels or job choice alternatives. Equation (158) is a matrix of criterion estimates and (159) is a matrix of observed criterion values. There is considerable discussion in the literature under the heading of "across-subjects vs. within-subjects analysis." This discussion appears to be based on a structuring of the data according to equations (158) and (159). This discussion itself seems to grow out of certain unconventional and irregular, if not questionable, approaches to model construction and data analysis that plague much of the work in expectancy theory. One wonders whether there has been too much reliance on traditional inappropriate statistical models that continue to reappear in elementary or even more advanced textbooks in statistics. The very expression "across vs. within analysis" smacks suspiciously of overworked models in experimental design and analysis of variance whose applications have far exceeded their relevance.

Limitation of the Within vs. Across Distinction

Certainly after obtaining the values in (158) according to the procedures implied in the foregoing discussion, one can find average row or column correlations between the matrices in (158) and (159). One can calculate and compare row and column means and standard deviations for the two matrices. Furthermore, one can even introduce constraints into the model so that row and column means and variance of the two matrices will be equal. But the model we have suggested dissolves the issue of across vs. within subject analysis by including both in one fell swoop of the computer.

Reducing the Number of Parameters

It might be noted that the number of parameters can be reduced if we want to assume that the levels within a dimension of the outcome modality shall have equal weight. Then within a dimension we would have unit weighting in the summation of the expectancy-valence products while we would retain the differential weights for the separate outcome dimensions. In the case of the two-level dimension model, this would cut the number of parameters in half. For the example given, instead of 49 parameters we would have only 25. If more levels were introduced, the number of parameters would not be affected by the simplified model, whereas it would be increased in the complete model.

LEVELS OF DIMENSIONS

This brings us to the subject of levels of dimensions and how they are established. As we have already seen, the homogeneous dimension modality presents no problems as, by definition, the dimensions are unilevel.

The Job Choice Model

A typical example in expectancy theory models is job valence used as a criterion. A list of possible jobs is provided the subject. He may be asked to indicate, on a ten-point scale, the overall satisfaction he feels or thinks he would feel in each of the positions listed. The scale may range from zero for "extremely unsatisfied" to 10 for "extremely satisfied." Such a scale does not have an absolute origin or a unit of measurement that may, in any sense, be considered rational.

This lack of absolute origin or rational unit is not necessarily a criticism of the scale. It should be noted, however, that such a 10-level scale does not apply to the job dimensions that have only one level but, rather, to the valence modality that has been accorded only one dimension but, in this example, has ten levels.

The Performance Level Model

In the previous example, as in most of the work in expectancy theory, the valent levels are assigned by the person to whom they apply. But we may cite an example where a criterion modality may not be self-assigned or be subjective in nature. Suppose the criterion is a unidimensional performance criterion such as in a card-punching job. The scale of productivity may range from less than 35,000 key strokes per day to 95,000 or more key strokes per day. There may be eight levels, including the extremes, proceeding by intervals of 15,000 strokes between the extreme intervals.

Here we do have a scale with an absolute origin and a rational unit of measurement. The number of levels is arbitrarily established at eight but the values assigned to the levels imply an absolute scale of measurement on which the levels are based.

We could now ask the card punchers to indicate, on a scale of 0 to 100, of what value it would be to them to perform at each of the eight levels. A rating of zero could be indicated as "of no value"; a rating of 50, as "moderately valuable"; and a rating of 100, as "extremely valuable." Such ratings would constitute the valences assigned to the eight levels of performance by the performers themselves.

But again the 100-level rating scale has neither an absolute origin nor a rational unit of measurement. Actually, to perform at less than 35,000 key strokes a day might be negatively valent to the worker. But, in this example also, the lack of an absolute origin or unit for the values assigned to the 100-level unidimensional valence modality is not necessarily objectionable, provided an adequate mathematical model and appropriate data analysis procedures are used.

Summary of Criterion Types

These two examples illustrate two quite different types of criterion modalities and the way levels of modalities may be treated. The first is a job choice model. It begins with a jobs modality of single level dimensions. A person assigns a level of a valence modality to each dimension. The values of the valence levels are arbitrarily assigned the number of 0 to 10.

The second is a performance level model. It begins with an objective performance scale segmented into eight levels. A valence dimension is arbitrarily assigned 100 levels and the numbers 0 to 100 are assigned to the successive levels. A person assigns a level of the valence modality to each of the eight performance levels.

The Contingency Modality

The contingency modality may be presented with a list of dimension statements about things that may impede or facilitate performance or production. In one experiment (Nebeker and Moy, 1976), presumably to simplify administration, 15 dimensions were presented without explicit levels. As a matter of fact, most of them could be interpreted as isolated levels from dimensions. For example, one statement is, "The documents are hard to read." This statement might well be regarded as a level from a dimension that could be called, "Legibility of documents." For this modality, one could specify levels of legibility. As we pointed out early in this document, levels of a dimension of a modality may themselves be treated as dimensions if there are more than two. There appears to be no coherent or comprehensive body of theory or practice on the use of levels as dimensions.

In any case, in this example, a scale is provided with the list of contingency dimensions and the subject is instructed to indicate how much each contingency dimension slows him down from doing as much as he might be able to do. The scale has 100 levels with numbers from 0 to 100 assigned to the successive levels. The lowest levels means, "Does not slow me at all," and the highest, "Slows me to a standstill."

The contingency modality is of particular interest because (1) it undoubtedly is of great importance in the overall productivity of a group of workers and (2) it has not been explicitly integrated into any mathematical model of expectancy theory. There is not even a generally accepted name for the construct that the contingency scale purports to measure, such as valence or expectancy.

The example cited does make some attempt to integrate the contingency modality with that of the unidimensional performance modality. The subject is instructed to indicate his expectancy of achieving each performance level for three different effort levels (top speed, 3/4 speed, 1/2 speed) under the assumption of a normal number of problems. However, there is no suggestion of how the subjects might relate their assignments of "slowdown" levels to the contingency dimension of a "normal number" of problems.

The scale for indicating expectancy of performance at each performance level is again a graphic rating scale of 100 levels from 0 to 100. The extremes are, "Can't do it" and "Can do it every time." Presumably this scale does, in theory at least, have a true origin and rational unit. The zero level does mean a zero probability and each level represents one percentage point. Whether or not subjects actually interpret the scale in this manner and to what extent there are individual differences in interpretation is another matter.

Units and Origin of Assigned Values

We have, then, examples of valence levels whose scale of assigned values does not have an absolute origin nor natural units. We also have performance levels whose scale of assigned values does have an absolute origin and natural units. We have an example of "slow down" levels for dimensions, but not levels of the contingency modality whose scale of assigned values appears to be comparable to an expectancy scale. This latter may be said to have both absolute origin and natural units, but we as yet have no overall mathematical model that explicitly incorporates contingency variables into an expectancy theory model. Finally, we have an example of expectancy levels whose scale of assigned levels does have an absolute origin and natural units, even though there might be questions concerning the interpretation of the scale by persons using it.

The Valence Modality

Next let us consider the values of valence levels assigned to a level of a dimension of an outcome modality. Since valences can be either positive or negative, a valence scale should presumably have an absolute origin for any specified level of an outcome dimension. Again, we may take an example from Nebeker (1976). The subject is presented with a list of outcomes (or levels of outcomes). For each, he is to indicate a valence value from a scale that goes by intervals of 1 from -10 through to +10. Minus ten is defined as, "I would do almost anything to avoid it," and +10 as, "I would do almost anything to get it." This scale has 20 valence levels. In another example from Nebeker (1976), a similar valence scale going from -5 through 0 to +5 has only 10 levels. In this scale -5 is interpreted as, "Extremely unhappy" and +5 as, "Extremely happy."

The important feature of both scales is that the values assigned to the valence levels assume an absolute zero. They do not, as in the case of expectancy scales, also assume a natural unit of measurement. It should be recalled that the valence scales for job dimensions and performance levels did not assume an absolute origin. Again, as in the case of the expectancy scale, we may raise the question of how persons will interpret the zero point of the scale. Will there be stable individual differences in the extent to which zero is actually interpreted as complete indifference or complete irrelevance to the behavior of the subject?

SCALE ORIGIN AND INTERACTION

This question does not concern us now. What does concern us is the relevance of the zero point for a central concept of expectancy theory models. This is the interaction of expectancy and valence in the influence of human behavior in general or work performance in particular. This interaction is expressed by the product of an expectancy value and a valence value. The summation of such values, weighted or unweighted, plays a central role, as we have seen, in predicting effort or performance in expectancy theory models.

The Schmidt Critique

Considerable discussion is found in the literature concerning the validity of this interaction variable. Schmidt (1973) has argued that, unless the scales used for measuring expectancy and valence have absolute zero points, the product terms are meaningless. Nebeker and Mitchell (1973) have responded in effect that the instructions and the formats of both the expectancy and valence scales can be adequate to guarantee reasonable confidence in the zero points of the scales. Their arguments are persuasive and the scales that Nebeker and Moy (1976) and their colleagues have recently provided for these variables would lead one to believe that constant errors in origin shift would not be expected. It is questionable, therefore, whether systematic research to test the assumption of an absolute zero in currently available formats would be justified on any large scale.

The Multiple Regression Alternative

Nebeker and Mitchell (1973) appear to be more concerned with the selection of the outcome modality dimensions than with the problem of the absolute zero of expectancy and valence scales. Actually, even this concern may lose some of its force with the use of the multiple regression model that we have suggested. This model would mainly require that the relevant outcomes were adequately covered by any list that might be generated.

Problems of levels of specificity-generality, overlapping, or even inclusion of irrelevant outcomes would not be very serious in an empirical multiple regression model with plenty of degrees of freedom for yielding stable regression weights. By stepwise or other available techniques, one could eliminate redundant or irrelevant items.

The Values of Levels

This discussion of levels of dimensions has brought to light a point that has not been previously explicitly recognized but that has been implicit in much of the discussion in this and earlier sections of this document. It is the distinction between levels of dimensions and the values assigned to these levels. In some modalities, the distinction is important. In others, it is not.

The Number of Levels

The modalities for which the distinction is most important are the expectancy and valence modalities. As we have seen, 100 plus 1 levels have been assigned expectancy modalities by some of the Nebeker formats but we must also specify, in such cases, that the first of these levels be assigned a value of zero and each succeeding one an increment of .01. Otherwise, we cannot logically use an assigned value as an expectancy in the expectancy-valence product because, however arrived at, expectancy is a person's estimate of the probability that an event will occur. Whatever the number of levels assigned the modality, it is important that the lowest level be assigned a value of zero and the highest a level of 1. The assigned interval values must increase by the reciprocal of the total number of levels minus one.

The total number of levels assigned the expectancy dimension is arbitrary. Perhaps the chief consideration is to have the number large enough so that no one may be expected to make reliable discriminations that are finer than those provided by the number of levels. A safe rule is that it is better to exceed the discriminatory ability of all members of a group than to provide for fewer choices than can be reliably used by the most discriminating members of the group.

This same principle in determining the number of levels provided is generally applicable to most situations where persons are required to give ratings of variable entities. It is as true with valences as with expectancies. However, while the restrictions are placed on the lower negative and upper positive limits of values assigned to the levels, it is important that the level assigned the zero value must be clearly defined as one of complete indifference. Since it is only the expectancy and valence values that necessarily appear as product terms in expectancy theory models, these are the modalities for which the assignment of values to levels is most crucial.

The Mathematics of the Scale and Origin Problem

We may, however, explore further the reservations of Schmidt (1973) with reference to the expectancy-valence product terms in expectancy theory models.

Definition of Notation

Suppose we let

$h k^{p_{ij}}$ = the expectancy of person h that the i th level of the prior dimension will be followed by the j th level of subsequent dimension k

$$\begin{aligned}
{}_k V_j & \text{ be the valence of level } j \text{ of dimension } k \\
n_i & \text{ be the number of levels of the prior dimension} \\
{}_k n_j & \text{ be the number of levels of dimension } k \\
n_k & \text{ be the number of subsequent modalities} \\
hk^\rho & \text{ be the } n_i \times {}_k n_j \text{ matrix of } hk^\rho_{ij} \text{'s} \\
{}_k V & \text{ be the } n_k \text{ th order vector of } {}_k V_j \text{'s} \\
{}_k F & = a_k^\rho {}_k V & (160) \\
F & = ({}_1 F, \dots, {}_{n_k} F) & (161) \\
\tilde{Y} & = F U & (162)
\end{aligned}$$

where U is a vector to be determined and \tilde{Y} is an estimate of the criterion.

The Mathematical Development

Now in (160) ${}_k V$ is a vector of valence level values assigned to the levels of dimension k . Suppose we assume that the values that were assigned to the valence levels had neither an absolute origin nor a natural unit so that we require a linear transformation of the scale of values to make it an absolute scale. We let V be from the arbitrary scale and v be from the absolute scale so that

$$v = A_V 1 + B_V V \quad (163)$$

where A_V and B_V are the origin and scaling constants, respectively.

Similarly, we assume an arbitrary scale for the expectancies that result in a matrix S from the arbitrary scale, so that

$$\rho = A_\rho 1 1' + B_\rho S \quad (164)$$

From (163) and (164)

$$\rho v = (A_\rho 1 1' + B_\rho S) (A_V 1 + B_V V) \quad (165)$$

From (165)

$$\rho v = A_\rho A_V n_i 1 + A_\rho B_V 1 1' V + B_\rho A_V S 1 + B_\rho B_V S V \quad (166)$$

Now let

$$C_1 = A_\rho A_V n_i + A_\rho B_V 1 1' V \quad (167)$$

$$C_2 = B_\rho A_V \quad (168)$$

$$C_3 = B_\rho B_V \quad (169)$$

From (167), (168), (169) in (166)

$$\rho v = C_1 I + C_2 S I + C_3 S V . \quad (170)$$

If we reintroduce the subscripts from (160) into (170), we may write

$${}_k F = {}_k C_1 I + {}_k C_2 {}_{hk} S I + {}_k C_3 {}_{hk} S {}_k V . \quad (171)$$

If we assume that the arbitrary scale constants for both expectancy and valence remain the same for all n_k dimensions, we may drop the k prescripts from the C 's in (171) and write

$${}_k F = C_1 I + C_2 {}_{hk} S I + C_3 {}_{hk} S {}_k V . \quad (172)$$

Then we let

$${}_{hk} S I = {}_{hk} L \quad (173)$$

$${}_{hk} S {}_k V = {}_{hk} M . \quad (174)$$

From (172), (173), and (174),

$${}_k F = C_1 I + C_2 {}_{hk} L + C_3 {}_{hk} M . \quad (175)$$

Now let

$${}_h L = ({}_{h1} L, \dots, {}_{hn_k} L) \quad (176)$$

$${}_h M = ({}_{h1} M, \dots, {}_{hn_k} M) . \quad (177)$$

From (160) to (162) and (175) to (177),

$${}_h \tilde{Y} = C_1 I' U I + C_2 {}_h L U + C_3 {}_h M U . \quad (178)$$

We redefine

$$C_1 = C_1 I' U . \quad (179)$$

So, from (178) and (179),

$${}_h \tilde{Y} = C_1 I + C_2 {}_h L U + C_3 {}_h M U . \quad (180)$$

Now, if n is the number of persons, we can let

$$\tilde{Y} = \begin{bmatrix} {}_1 \tilde{Y} \\ \vdots \\ {}_n \tilde{Y} \end{bmatrix} \quad (181)$$

$${}_h L = \begin{bmatrix} {}_1 L \\ \vdots \\ {}_n L \end{bmatrix} \quad (182)$$

$$h^M = \begin{bmatrix} 1^M \\ \vdots \\ n^M \end{bmatrix} \quad (183)$$

$$K = \begin{bmatrix} 1^C & 1 \\ \vdots & \\ n^C & 1 \end{bmatrix} \quad (184)$$

and from (180) through (183) we can write

$$\tilde{Y} = K + C_2 L U + C_3 M U. \quad (185)$$

The Case of Unit Weights

A common assumption in expectancy theory models is that the outcome dimensions should have equal weight. This means, without loss of generality, the vector U in (185) is a unit vector. If we let

$$W_2 = L 1 \quad (186)$$

$$W_3 = M 1 \quad (187)$$

we can, with the assumption of unit weights, write

$$\tilde{Y} = K + C_2 W_2 + C_3 W_3. \quad (188)$$

But let us look more closely at equation (178) and see what happens if we assume that the subsequent k dimensions all have two levels and that they are conditional rather than joint expectancies. In this case, the sum of the two expectancies corresponding to a given level of the prior dimension would be 1. Therefore

$$h_k^S 1 = 1. \quad (189)$$

If we let

$$k^{C_0} = k^{C_1} + k^{C_2} \quad (190)$$

$$k^C = k^{C_3}, \quad (191)$$

then, from (178) and (189) through (191),

$$k^F = k^{C_0} 1 + k^C h_k^S k^V. \quad (192)$$

Equation (192) will apply even if the number of levels is greater than two so long as the expectancies for all levels of an outcome dimension corresponding to a given level of the prior dimension sum to 1. Therefore, under this constraint of the expectancies, we can write

$$\tilde{Y} = (K + C_2 I) + C_2 W_3 . \quad (193)$$

Let us also put the subvectors of \tilde{Y} as rows and the same for the right side of (188). Then we can write

$$\tilde{y} = D_K I I' + C_3 w \quad (194)$$

where

$$D_K = C_1 + C_2 \quad (195)$$

and w is a matrix whose columns are the subvectors of W_3 . Suppose we also let y be a matrix constructed from the criterion supervector Y . Then we write the residual matrix

$$y - (D_K I I' + C_3 w) = \epsilon . \quad (196)$$

Suppose now we let

$$Z = y \left(I - \frac{1 I'}{n_k} \right) \quad (197)$$

$$X = w \left(I - \frac{1 I'}{n_k} \right) \quad (198)$$

$$e = \epsilon \left(I - \frac{1 I'}{n_k} \right) . \quad (199)$$

From (196) through (199)

$$Z - C_3 X = e . \quad (200)$$

Constant Transformation Parameters

Now, under the assumptions of (1) linear transformations of both the expectancy and valence level values that remain constant for both persons and outcomes, (2) unit weight of the outcome dimensions, and (3) conditional rather than joint expectancies, equation (200) holds. This equation says that, under the above assumptions, if (1) the actual and estimated criterion vectors are reordered into person-by-level or choice matrix, and (2) both matrices are deviated by rows, then the correlation between corresponding elements of the deviated matrices will be independent of the transformation constants for the expectancy and valence scales.

This conclusion follows from equation (200) because it was derived from the given assumptions and only one unknown constant, C_3 , remains in the equation. Obviously, irrespective of this constant, the correlation between corresponding elements of Z and X are the same. We see therefore that, even under the fairly liberal assumptions, the objections of Schmidt (1973) are not cause of serious concern.

Variable Transformation Parameters

We may, however, relax the assumption of constant transformations by persons and assume that there may be individual differences in the interpretation of both the expectancy and valence scales. We may require both an additive and a scaling constant for each person for each of the two scales. We shall therefore return to equation (172). First we put the prescripts h on the C 's and on ${}_k^F$ and ${}_k^V$ to show that they apply to a single person, thus:

$${}_{hk}^F = {}_h^C{}_1 + {}_h^C{}_2 {}_{hk}^S{}_1 + {}_h^C{}_3 {}_{hk}^S {}_{hk}^V. \quad (201)$$

If we assume the S 's to be constrained to conditional expectancies, we have

$${}_{hk}^S{}_1 = 1. \quad (202)$$

We let

$${}_h^C{}_0 = {}_h^C{}_1 + {}_h^C{}_2 \quad (203)$$

$${}_{hk}^S {}_{hk}^V = {}_{hk}^M. \quad (204)$$

Substituting (202), (203), and (204) in (201) gives

$${}_{hk}^F = {}_h^C{}_0{}_1 + {}_h^C{}_3 {}_{hk}^M. \quad (205)$$

Suppose now we sum (205) over the k outcomes to get a criterion estimate ${}_h^{\sim}Y$ for the h th person. We let

$${}_h^W = \sum_{k=1}^{n_k} {}_{hk}^M \quad (206)$$

$${}_h^B = n_k {}_h^C{}_0 \quad (207)$$

$${}_h^D = {}_h^C{}_3 \quad (208)$$

$${}_h^{\sim}Y = \sum_{k=1}^{n_k} {}_{hk}^F. \quad (209)$$

From (205) through (209)

$${}_h^{\sim}Y = {}_h^B{}_1 + {}_h^D {}_h^W. \quad (210)$$

Suppose now we write

$$\tilde{Y} = \begin{bmatrix} {}_1^{\sim}Y \\ \vdots \\ {}_n^{\sim}Y \end{bmatrix} \quad (211)$$

$$W = \begin{bmatrix} 1 & W' \\ \vdots & \vdots \\ n & W' \end{bmatrix} \quad (212)$$

$$B = \begin{bmatrix} 1 & B \\ \vdots & \vdots \\ n & B \end{bmatrix} \quad (213)$$

$$D = \begin{bmatrix} 1^D & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & n^D \end{bmatrix} \quad (214)$$

From (210) through (214)

$$\tilde{Y} = B I' + D W . \quad (215)$$

Suppose now we also assume the criterion matrix Y of order $n \times n_i$ where the rows are persons and the columns are levels or choices in the criterion variable. We might also assume that, since persons may differ in their interpretation of the valence scale, each one should have his criterion values adjusted by an additive and a scaling constant. We might then write the adjusted criterion matrix as

$$Y_2 = B_2 I' + D_2 Y . \quad (216)$$

In order to generalize the forms of (215) and (216), we shall rewrite them, respectively, as

$$Y_1 = B_1 I' + D_1 W_1 \quad (217)$$

$$Y_2 = B_2 I' + D_2 W_2 \quad (218)$$

Now the matrices W_1 and W_2 are given and the B 's and D 's presumably are to be determined so that Y_1 and Y_2 are as nearly the same as possible, according to some defined loss function.

A number of rationales for the solution of this problem have been considered at length and will be taken up in more detail later. As an abstract mathematical model, the problem is of considerable interest. One of the interesting aspects of the problem concerns the formulation of constraining equations or conditions to be imposed on the solution for the B and D parameters. It is obvious from equations (217) and (218) that, if $D_1 = D_2 = 0$ and $B_1 = B_2$, the corresponding elements of Y_1 and Y_2 would be identical. The solution would, of course be trivial.

Modeling for Scale and Origin Parameters

Models may be developed in which the parameters to be solved for are scale and origin parameters for the dimensions of the person modality. Examples of such modeling

are given in Appendices A through H. These models vary essentially in the way particular constraints are imposed for the solution of the scaling and origin parameters.

Linearity or Nonlinearity

One issue that continues to remain blurred in the fog of expectancy theory literature is that of the validity of the expectancy-valence product terms, as well as the use of other product, as distinguished from linear functions. The debate is fairly pointless and arises from a misconception of scientific methodology. The notion that one should concentrate on small dismembered bits and pieces of an integrated theory is not restricted to Campbell and Pritchard's (1970) criticism of the grand design approach to expectancy research. Rather, it characterizes the attitudes of those ruggedly individualistic free-enterprising social scientists who constitute a large proportion of the behavioral sciences community.

One does not need to agonize over the question of whether linear or nonlinear functions should be used in the model. Campbell's (1976) caveat that "to propose explicit multiplicative combinations or other higher order functions is going a bit too far beyond our present measurement capability" is itself going a bit too far in view of the ease with which both linear and interactive functions can be included—and tested—in mathematical models.

To illustrate this point, let us return to equation (201). This equation is linear in the expectancies and, as can be seen from equation (167), in the valences. The nonlinear interaction products of expectancies and valences are also included in the last term on the right of equation (201). True enough, this equation together with the further complication of the model suggested by equation (150), implies the incorporation of a large number of parameters. However, if the mathematical model is clearly formulated, the parameters clearly specified, the loss and optimizing functions properly formulated, the constraining equations well conceived, the significance test appropriately chosen (in short, if the functional component of the multivariate analysis system is correctly fashioned and executed), the issue of linearity vs. nonlinearity becomes irrelevant.

Quantification of Concepts

Quantification technology should play a fundamental role in all behavioral science investigations, and expectancy theory research is no exception. It may be said that check lists, rating formats, and instructions to the subjects provide the blueprint and specifications for the multivariate expectancy model. Concepts may generate check lists but, once these check lists have been generated, the quantification formats designed, and the instructions drafted, they define the concept: they are in fact the concept. If any one denies that they are, the burden of proof is on the challenger to prepare his own list, his own formats, and his own instructions. These and these only are acceptable as his definition of the concept. We suggest that, until concepts are operationalized in terms of the values of specified levels of verbalized dimensions of a modality, until formats and instructions are provided as to how and what evaluator assigns values of the levels to actual entities, until such operationalization takes place, the concept can have no more than literary merit.

THE THREE-MODALITY MODEL

So far we have considered primarily the original Vroom models and modifications of them by Nebeker and his associates. One reason for this restricted treatment is that

other proposed models have been less explicit in their specification of the variables involved and in exact mathematical formulations for the models. Our mathematical developments have included primarily the performance and outcome modalities with the possible telescoping of needs with the latter.

A more complete treatment would also include the effort modality, together with a specification of effort levels. A model including effort prior to performance should provide for the elicitation of either joint or conditional expectancies for each of its levels, followed by each of the performance levels. It should also provide for the elicitation of valence for each of its levels. In addition, it might also provide for the elicitation of expectancies corresponding to each effort level, followed by each level of each dimension of the outcome modality. There is no logical reason why joint or conditional expectancies for every level of a dimension of a prior modality, followed by every level of every dimension of any or all subsequent modalities, should not be elicited. The practical difficulties of responding to all such possibilities are obvious. The complexities of incorporating such data into a mathematical model are forbidding. So far, it appears that only expectancies from contiguous modalities have been utilized in expectancy models and it would probably not be profitable to extend the model beyond contiguous modality expectancies.

The Galbraith-Cummings Model

In any case, we might attempt to see how the formulation of a three-modality model might be approached. This model consists of the three modality links: effort, performance, outcomes. The first two of these are treated as unidimensional and the third, as multidimensional. Galbraith and Cummings (1967) give the equation

$$W = E \left(\sum_{j=1}^n I_{ij} V_j \right) . \quad (219)$$

They define the symbols as follows:

W = effort

E = expectancy that effort leads to performance

I_{ij} = the instrumentality of performance for the attainment of second level outcomes

V_j = the valence of the second level outcome

n = the number of outcomes

Modification of the Model

We should first like to object to the term "second level outcome" for a number of reasons:

1. It conflicts with our use of the term level of a dimension. Priority of usage might be claimed as justification for its use.

2. The designation reflects the unfortunate failure to generalize to a logical or chronological sequence of modalities required by an adequate expectancy theory model.

3. This usage of "level" calls attention to the conspicuous failure of many expectancy model theorists to recognize the need for the concept of levels of dimensions of modalities.

We must also object to the definition of E because it ignores both levels of effort and levels of performance. We have already discussed at length the concept of instrumentality (I_{ij}) and concluded that, whatever instrumentality can do, expectancy can do better. The definition of V_j as a valence of an outcome again fails to recognize levels of an outcome. The least one might do is to specify a two-level (0, 1) outcome as Nebeker et al. have done.

Definition of Notation

Let us see now whether we can modify equation (219) into a form more in line with the foregoing discussions. We begin with a definition of notation. We let

- n_i be the number of levels in the effort modality
- n_j be the number of levels in the performance modality
- ${}_d n_k$ be the number of levels in dimension d of the outcome modality k
- n_d be the number of outcome modalities
- P_{ij} be the expectancy that level i of the effort dimension will be followed by level j of the performance dimension
- ${}_d P_{jk}$ be the probability that level j of the performance dimension will be followed by level k of dimension d of the outcome modality
- ${}_d V_k$ be the valence of level k of dimension d of the outcome modality
- V_j be the valence of performance level j
- W_i be the valence of effort level i

The Scalar Algebra Formulation

We assume that the valence of effort level i is

$$W_i = \sum_{j=1}^{n_j} (P_{ij} V_j) \quad (220)$$

and that

$$V_j = \sum_{d=1}^{n_d} \left(\sum_{k=1}{{}_d n_k} ({}_d P_{jk} {}_d V_k) \right) . \quad (221)$$

From (219), (220), and (221),

$$W_i = \sum_{j=1}^{n_j} P_{ij} \left(\sum_{d=1}^{n_d} \left(\sum_{k=1}{{}_d n_k} ({}_d P_{jk} {}_d V_k) \right) \right) . \quad (222)$$

Presumably now we might use (222) to estimate the valence of any specified level of effort and the highest W_i might be taken as an estimate of the actual effort level chosen. The expectancy and valence values might be determined by means of rating formats and instructions already discussed and used by previous investigators. If only two-level outcome dimensions were used, equation (222) could be simplified to

$$W_i = \sum_{j=1}^{n_j} P_{ij} \left(\sum_{d=1}^{n_d} ({}_dP_j {}_dV + (1 - {}_dP_j) {}_d\bar{V}) \right) \quad (223)$$

where ${}_dV$ and ${}_d\bar{V}$ in the inside parentheses on the right of (223) mean respectively the valence of level 1 of dimension d of the outcome modality and level 0 (nonoccurrence) of this dimension.

It should be noted that (222) and (223) do not explicitly include intrinsic valences of the levels of either the effort or performance dimensions. These may well be important in the estimation of the overall valence of each effort level.

Another significant feature of (222) and (223) is that, like most current expectancy theory models, they do not include parameters to be solved for. This failure to achieve the status of a full-fledged multivariate analysis model could be corrected with a development much along the same lines we used with the modified Vroom model. In this development, we shall assume that the weights to be assigned the individual product terms in the inside parentheses on the right of equation (222) are all unity, as implied by the inner summation sign. We shall assume, however, that the dimensions of the modalities may not necessarily have equal weight in estimating the valences of the levels of the performance dimension. We let U_d be the weight of the d th dimension of the outcome modality. Then equation (222) may be rewritten

$$W_i = \sum_{j=1}^{n_j} P_{ij} \left(\sum_{d=1}^{n_d} U_d \sum_{k=1}^{n_k} ({}_dP_{jk} {}_dV_k) \right) \quad (224)$$

The Matrix Algebra Formulation

To convert from the clumsy scalar notation to the simpler and more elegant matrix notation, we first let

${}_dP$ be the $n_j \times n_k$ matrix of the ${}_dP_{jk}$ expectancies

${}_dV$ be the vector of order n_k of the ${}_dV_k$ valences

b^V be a vector of order n_j

U be a vector of order n_d of the weights W_d

$$b^V = {}_dP {}_dV \quad (225)$$

$$v = ({}_1V, \dots, {}_{n_d}V) \quad (226)$$

$${}_jV = v U \quad (227)$$

Now jV in (227) is an n_j th order vector of the V_i in equation (221).

Next we let

P be an $n_i \times n_j$ matrix of the P_{ij} expectancies

W be a vector of order n_i of the estimated valences of the n_i effort levels.

Then, from the definitions and equations (224) through (227),

$$W = P v U \quad (228)$$

If we let

$$P v = G \quad (229)$$

we have, from (228) and (229)

$$W = G U . \quad (230)$$

To summarize the steps for estimating the vector of estimated valences for the effort levels, we have

$$bV = d^P dV \quad (231)$$

$$v = ({}_1V, \dots, {}_kV, \dots, {}_{n_d}V) \quad (232)$$

$$G = P v \quad (233)$$

$$W = G U . \quad (234)$$

The General Case

The development so far has considered only a single person. Suppose therefore we rewrite equation (234) with prescripts h and, instead of W , we use \tilde{Y} to indicate an estimate of the criterion Y . We have then from (234)

$${}_h\tilde{Y} = {}_hG U \quad (235)$$

We note that U does not carry the prescript h since it is assumed the same for all cases.

Next we let

$$\tilde{Y} = \begin{bmatrix} {}_1\tilde{Y} \\ \vdots \\ {}_h\tilde{Y} \\ \vdots \\ {}_n\tilde{Y} \end{bmatrix} \quad (236)$$

$$G = \begin{bmatrix} {}_1G \\ \vdots \\ {}_hG \\ \vdots \\ {}_nG \end{bmatrix} \quad (237)$$

where n is the number of persons. Then from (235), (236), and (237),

$$\tilde{Y} = G U . \quad (238)$$

Suppose now we have a vector Y of ratings by supervisors or other acceptable evaluators of the actual effort levels normally exerted by each person. The evaluator might indicate with a check mark the applicable level for each person. This level would then be assigned a value of 1. The other levels would all be assigned values of zero, or the evaluator might indicate, for each effort level, the proportion of time the person exerted that level of effort. In any case, we have here an example for which the person assigning the value and the one for whom it is assigned are not the same.

Returning to equation (238), we could write the residual equation

$$(Y - G U) = \epsilon \quad (239)$$

where Y is the criterion, G the predictors, and U is the vector of parameters to be solved for. Again, we have the classical multiple regression form where the solution for U is well known. If desired, an additive constant can be introduced as a constraint on U so that the means and sigmas of Y and \tilde{Y} would be equal. Again also, the person subvectors of Y and \tilde{Y} could be ordered into two $n \times n_i$ matrices so that the rows are persons and the columns effort levels. One could then conduct both within- and across-subject analyses, whatever that means. Actually the procedure as outlined has included both within- and across-subject data.

INTERVENTION VARIABLES

We have seen how a triple-link modality chain may be incorporated into an expectancy model both with and without parameters. This chain consists of the effort, performance, and outcome modalities. Between the first two, we have the matrix of joint or conditional expectancies. Between the unidimensional performance modality and the multidimensional outcome modality, we have an expectancy matrix for each dimension of the outcome modality. In general, we might expect that the higher the correlations between the valences and expectancies, the higher will be predicted levels of effort.

The Contingency Variables

This brings us to a consideration of how these expectancies might be increased. First, let us consider the expectancies between effort and production. We have seen that the contingency modality involves dimensions that may influence productivity or performance. Some of these contingency variables may be beyond the control of the worker but subject to the control of management. These are variables that we referred to earlier as manipulable or intervention variables. They can be of great importance for the overall objectives of an organization. The values of the levels of dimensions may presumably be manipulated by management so that worker effort will be more productive. Such manipulation can therefore increase the expectancies that higher levels of effort will be followed by higher levels of performance or production. In this way, by the manipulation of contingency variables, the correlations of the expectancies with performance levels may be increased.

The Expectancy Variables

The situation with respect to the matrices of expectancies is somewhat different. Management may manipulate the contingency variables so as to modify indirectly the expectancies between effort and performance. It may also directly modify the objective probabilities between performance and the dimensions of the outcome modality. These objective probabilities may, in turn, modify expectancies (subjective probabilities) and thus also their correlations with the valences of outcome levels. As these correlations increase, the valences of higher performance levels should also increase and these higher valences should be followed by higher effort valences.

Summary of Intervention Variables

We see then how the manipulation of contingency variables and objective probability variables between production and outcomes could result in higher levels of productivity. This latter source for stimulating productivity has been recognized from the first by expectancy theorists working in organizational settings. The manipulation of contingency variables has long been recognized by management as a technique for increasing productivity. However, expectancy theorists have not yet explicitly incorporated contingency variables into a mathematical formulation of an expectancy theory model.

THE POSTULATES OF BEHLING AND STARKE

Introduction to the Postulates

We have referred throughout this document to various contributors to expectancy theory literature from the time of Vroom to the present. One contribution that merits rather detailed consideration is by Behling and Starke (1973), entitled "The Postulates of Expectancy Theory." The paper, in the authors' own words

1. Describes briefly a basic expectancy model and indicates how it relates to other formulations.
2. Identifies, in the form of postulates, the assumptions underlying the basic and elaborated forms of expectancy theory, as well as the state of empirical testing of the postulates.
3. Draws some conclusions about the validity of expectancy theory and about the direction in which research regarding it might profitably move.

The first part of the paper begins with Vroom's original force model and indicates various modifications that have been introduced by later investigators. The modifications appear to have been largely on the verbal level rather than in a formal mathematical equation from which the functional component of a multivariate analysis system was developed.

The second part of the paper will be considered in more detail, since it lists specific postulates, the examination of which may contribute to a further development of a more comprehensive expectancy theory model. In introducing their postulates of expectancy theory, the authors include a paragraph that they appear to regard as important. We shall therefore quote it verbatim, itemizing the statements in the paragraph. For the most part, the statements appear to be literary productions. We shall attempt to see to what

extent they can be translated into an organismic-environmental terminology. The statements of the paragraph are as follows:

1. Expectancy theory is descriptive rather than prescriptive.
2. Prescriptive theories of behavior attempt to state the manner in which individuals should act in order to maximize utility satisfaction or some other variable.
3. Thus, they are properly judged on the basis of whether or not the application of the principles they contain increases the likelihood of maximizing favorable outcomes.
4. Whether the individuals actually behave in accordance with the principles is of secondary importance.
5. Descriptive theorists, on the other hand, purport to describe how individuals actually do act.
6. Such theories must be judged on their accuracy as reflections of behavior.
7. Whether the behavior leads to desirable outcomes is of secondary importance.
8. Unless expectancy theory and the assumptions that underlie it are realistic appraisals of individual behavior, the descriptive usefulness of expectancy theory as a whole must be severely questioned.

To determine what, if anything, these statements mean, let us first rewrite them stripped of excess verbage:

1. Expectancy theory is descriptive rather than prescriptive.
2. A prescriptive theory states how persons act to maximize valence.
3. They are judged on the extent to which the use of their principles maximizes valent outcomes.
4. Whether persons act in accordance with the principles is not important.
5. Descriptive theories describe how persons act.
6. They are judged on how accurately they reflect acts.
7. Whether the acts are followed by valent outcomes is not important.
8. The assumptions of expectancy theory, to be acceptable, must correspond closely to the acts of persons.

Next let us note that statements 4 and 7 appear to be two different ways of saying the same thing. They both seem to say that, for both types of theories, it is not important that acts actually performed are followed by valent outcomes. If these two statements do not serve to distinguish prescriptive and descriptive theories, we shall eliminate them and proceed to a further simplification of the remaining five statements by making numbers 2 and 3 prescriptive theories and numbers 5, 6, and 8 descriptive theories:

2. What acts maximize valence?
3. Do these acts maximize valence?
5. What acts do persons choose?
6. Do persons choose these acts?
8. Can the choice of acts be predicted?

Obviously, statements 2 and 3 are slightly different ways of asking the same question, as are statements 5 and 6. Therefore, we can drop the second of each pair. This leaves us, one prescriptive theory (No. 2) and two descriptive theories (Nos. 5. and 8).

Except for the joy of semantic jousting, probably no one would deny that the basic assumption of expectancy theory is as follows: Those acts are chosen that have the highest expectancy of leading to outcomes of greatest valence. That is what expectancy theory is all about. That is the basis of predicting acts. One might argue that the essential difference between statements 2 and 5 is that 2 implies a high objective probability that certain acts will be followed by positively valent outcomes, whereas 5 implies a high subjective probability that ~~certain~~ acts will be followed by positively valent outcomes.

It is doubtful that the authors meant to say that prescriptive theories are concerned with the objective probability that a positively valent outcome will follow an act, whereas descriptive theories are concerned with the subjective probability of the occurrence of the sequence of events. If that is what they meant to say, there is a much shorter, simpler, more concise way of saying it, but it is probably not true. If there are more complex, subtle, and important differences between the two types of theories, they do not stand out clearly in the authors' well wrought literary formulations.

The Postulates

We shall therefore assume that the paragraph we have discussed has no great significance for a comprehensive expectancy theory model and proceed to an examination of the postulates presented by the authors.

Comparability

Postulate 1. "Comparability. Any two $\Sigma (E_{ij} V_j)$'s are directly comparable. Either $\Sigma (E_{ij} V_j)_1$ is preferred to $\Sigma (E_{ij} V_j)_2$, $\Sigma (E_{ij} V_j)_2$ is preferred to $\Sigma (E_{ij} V_j)_1$, or the individual is indifferent between the two."

First we suggest that the postulate as stated is not meaningful. This is because the expression $\Sigma (E_{ij} V_j)$ is not defined. Presumably, the external subscripts 1 and 2 are meant to refer to the subscript i on the symbol E that refers to act i. The expression

itself is, according to Vroom and others, supposed to yield what we have called the disembodied spirit i and what they call the "force to perform" act i . What is meant by "the force to perform act 1 is preferred, not preferred, or equally preferred to the force to perform act 2" is far from clear. As we have suggested in our discussion of Vroom's behavioral choice model, the F_i to which the product summation is equated should be replaced by some symbol such as \tilde{V}_i to represent the estimated valence of act i , some other direct evaluation of act i such as performance measure, or a choice specification from a number of dimensions of a modality (jobs).

But this estimate derived from the product summation is actually a specific number. The E_{ij} 's and V_j 's are presumably provided by a person in terms of actual numbers, according to specific instructions and a specific quantification format. Therefore, the product summations are also specific numbers (e.g., 9 and 12 for $i = 1$ and 2 respectively). It probably would be meaningless to ask a person whether he preferred the number 9 to the number 12, or vice versa, or whether he was indifferent between the two. It could, of course, make a great deal of sense to a person to ask whether he preferred job 1 to job 2, or was indifferent, but that is quite another matter. It would also make sense to ask whether he preferred outcome j to outcome k , whose valences are represented by the V_j 's within the parentheses of the product summations.

In view of the fact that postulate 1 is difficult if not impossible to interpret, it is not surprising that the authors conclude that studies purporting to verify it have not been notably successful. The authors close their discussion of the postulate with "the assumption of unidimensionality of $\Sigma (E_{ij} V_j)$'s is open to question."

The problem with this conclusion is at least two-fold.

1. It does not make explicit that the comparisons to be meaningful must actually be among some defined species of valences of the acts 1, 2, and 3, rather than purely mechanical sets of calculations involving the E_{ij} 's and V_j 's.

2. Assuming that events or entities were defined so that expressions of preference could be elicited, the conclusion implies that comparisons may be based on a multidimensional, rather than a unidimensional, continuum. It is indeed likely that unidimensional models of the valence modality, as well as of the effort and performance modalities, are unrealistic oversimplifications of behavior phenomena. But this should not deter us from building multidimensionalities into an expectancy theory model rather than abandoning the model completely. Vast bodies of theoretical and experimental material are available in the factor analytic and multidimensional scaling fields that could contribute greatly toward extending the model to multidimensional modalities. This brings us to the second postulate.

Transitivity

Postulate 2. "Transitivity. Preference ordering of $\Sigma (E_{ij} V_j)$'s and of V 's are transitive. Specifically,

- a. Given $\Sigma (E_{ij} V_j)_1 > \Sigma (E_{ij} V_j)_2$ and $\Sigma (E_{ij} V_j)_2 > \Sigma (E_{ij} V_j)_3$, then $\Sigma (E_{ij} V_j)_1 > \Sigma (E_{ij} V_j)_3$, where $>$ is read as "having greater motivational force than."

b. Given $V_1 > V_2$, and $V_2 > V_3$, then $V_1 > V_3$, where $>$ is read as "greater than."

We can probably dismiss the inequalities in (a) above as obviously true since the summations are calculated numbers. Therefore, this part of the postulate is irrelevant. The second part (b), implies that, experimentally, the valence of each outcome is compared with all other outcomes. This is the classical method of paired comparisons on which Thurstone (1927) based his law of comparative judgment and developed his well known scaling technique. His procedure required that the paired judgments be made by a number of persons. Horst (1932) extended the technique to include also the pairing of positively and negatively valent outcomes as a basis for determining an absolute zero valence. Neither the Thurstone model nor Horst's extension of it attempted a single case solution for valences on the basis of paired stimulus elements.

Although there are problems with part (b) of postulate 2, we may examine some interesting implications of the inequalities given here for the single case. We note first that, if we attempt to apply part (b) to the expectancy model, the V_j , as such, can have meaning only if they apply to modalities with homogeneous dimensions (single-level dimensions) or to levels of a dimension of a modality. As we have seen earlier, the expectancy theorists and their critics have been sadly remiss in recognizing that the dimensions of the outcome modality are, in general, multilevel. To speak of the valence of an outcome modality without specifying a specific level is at least ambiguous if not meaningless.

In any case, let us take the paired comparison model in the classical Thurstonian sense and see how we may deduce valences for the outcomes on the basis of judgments of "greater than" or "less than." Suppose we start with a group of m stimulus elements with corresponding valences (V_1, \dots, V_n) for a given person. We assume that the subject judges all distinct pairs with no transivities. This means that they can be ordered so that

$$V_j - V_k > 0 \quad (240)$$

where $k > j$.

Suppose now we let

$$V_j - V_k = c + \epsilon_{jk} \quad (241)$$

where c is positive. From (241)

$$V_j - V_k - c = \epsilon_{jk} \quad (242)$$

We now have a multivariate analysis model represented by equation (242). The parameters of the model are the valences, V_j , and we shall take as the loss function

$$\Phi = \sum \epsilon_{jk}^2, \quad (243)$$

which is the classical least square function.

Before proceeding further, we shall first cast equation (242) in matrix notation. For simplicity, we take the case of $m = 4$. We let

$$X \equiv \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \quad (244)$$

$$V' \equiv (V_1, V_2, V_3, V_4) \quad (245)$$

$$I' \equiv (1, 1, 1, 1, 1, 1) \quad (246)$$

$$\epsilon' \equiv (\epsilon_{12}, \epsilon_{13}, \epsilon_{14}, \epsilon_{23}, \epsilon_{24}, \epsilon_{34}) \quad (247)$$

From (242), and (244) through (247),

$$X V - I c = \epsilon. \quad (248)$$

The optimizing function is from (243):

$$\Phi = \epsilon' \epsilon. \quad (249)$$

From (248) and (249),

$$\Phi = V' X' X V - 2 V' X' I c + I' I \quad (250)$$

For Φ to be an optimum (in this case, a minimum), we set the symbolic derivative of Φ with respect to the parameters equal to zero, thus:

$$\frac{\partial \Phi}{\partial V'} = 0. \quad (251)$$

But it can be shown from (250) that:

$$\frac{\partial \Phi}{\partial V'} = X' X V - X' I c. \quad (252)$$

Now from (244) it can be shown that

$$X' X \equiv \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix} \quad (253)$$

and

$$X' 1 c \equiv \begin{bmatrix} 4 \\ 1 \\ -1 \\ -3 \end{bmatrix} c . \quad (254)$$

More generally, for the case of m entities, we state, without going through the pedestrian arithmetic, that

$$X' X = m I - 1 1' \quad (255)$$

$$X' 1 = (n + 1) 1 - 2 U \quad (256)$$

where

$$U' \equiv (1, 2, \dots, m) . \quad (257)$$

Then from (252), (255), (256), and (257),

$$(m I - 1 1') V = ((m + 1) - 2 U) c . \quad (258)$$

From (258),

$$V = \frac{c (1 (m + 1 1' V) - 2 U)}{m} . \quad (259)$$

Now the model yielded by the paired comparison technique is independent of any arbitrary constant that might be added to or subtracted from the V_j 's. We may let $c = m / 2$ and write from (259)

$$V = 1 k - U \quad (260)$$

where k is a constant that is a function of any arbitrary origin. But k may be chosen so that the smallest V (i.e., V_m in (260)) is 1. This means from (257) and (260)

$$1 = k - m \quad (261)$$

or

$$k = m + 1 . \quad (262)$$

Then from (260), (261), and (262),

$$V' \equiv (m, m - 1, \dots, 1) . \quad (263)$$

Equation (263) shows that the least square solution to a perfectly consistent paired comparison model for a single case orders the entities on an ordinal scale. The constants c and k can be chosen to give this result. Any other constants would yield linear transformation of an ordinal scale.

It should be noted that, if intransitivities occur in the responses, a least square solution is still possible. Furthermore, it is not necessary to have prior information about

the ordering of the V's. For any pair to which the response is "less than," a -c would be substituted for a + c in (242). The matrix equation in (248) would be rewritten

$$X V - t c = \epsilon \quad (264)$$

where t is a vector of +1 and -1, according to whether the paired judgment is "greater than" or "less than." In general, where differences in valences were smaller, one could expect greater inconsistency and the solution could reflect these smaller differences in the solution for the V's.

One could, of course, provide a scale so that the subject could respond for each pair with, for example, one of the following: "much greater," "greater," "equal," "less," and "much less." These could be quantified respectively as -2, -1, 0, 1, 2. Then the t vector in (264) would consist of the values assigned from this scale. Such a modification should result in a better approximation to the subject's perceptions of the valences.

A much simpler approach, and one which has not been shown to be inferior to the paired comparison method, consists simply of rating formats that we have already discussed. The subject records his valence ratings of the dimensions or levels under consideration directly on a rating scale provided with specific and detailed instructions. Then, except for ties, the transitivity laws must hold by definition. As a matter of fact, the (b) part of postulate 2 should have been elaborated to take care of ties, thus: Given $V_1 \geq V_2$, $V_2 \geq V_3$, then $V_1 \geq V_3$.

The authors conclude with: "The implications of (intransitivities) for expectancy theory are negative, since intransitivities per se are detrimental to any maximizing theory." As we have indicated, intransitivities, if one insists on going about valence estimations the hard way, can actually bring what would otherwise be ordinal measures closer to the requirements of interval measures. Furthermore, the quotation appears to involve a nonsequitur. To say that intransitivities are detrimental to any maximizing theory seems to go far beyond what the current crude mathematical and pseudomathematical formulations would enable us to verify. We therefore suggest that no clear relevance of postulate 2 to expectancy theory has been established.

We shall now look at postulate 3 and see whether it may fare better than the first two.

Independence

Postulate 3. "Independence. There is no relationship between the valence of an outcome and the individual's estimate of the likelihood that it will follow from working at a certain level of effort. Specifically, $r_{E_{ij}, V_j} = 0$."

It should not be difficult to examine this postulate in view of the concise mathematical equation with which it concludes. We start with the well known formula

$$r_{XY} = \frac{\frac{\sum XY}{n} - \frac{\sum X}{n} \frac{\sum Y}{n}}{\sigma_X \sigma_Y} \quad (265)$$

Assume that, as in the postulate,

$$r_{XY} = 0 \quad (266)$$

From (265) and (266), we readily find that

$$\sum X Y = \sum X \frac{\sum Y}{n} . \quad (267)$$

Now in the case of postulate 3, we have

$$E_{ij} = X \quad (268)$$

$$V_j = Y . \quad (269)$$

Substituting (268) and (269) in (267) gives

$$\sum E_{ij} V_j = \sum E_{ij} \frac{\sum V_j}{n} . \quad (270)$$

Now presumably we are speaking of level i of effort and are summing over the expectancies that each of the levels of performance will follow. This means that the expectancies are conditional rather than joint and hence exhaustive. Therefore

$$\sum E_{ij} = 1 . \quad (271)$$

If n is the number of levels of performance and \bar{V} the mean of the performance level valences, we have

$$\bar{V} = \frac{\sum V_j}{n} . \quad (272)$$

From (270), (271), and (272), we get

$$\sum E_{ij} V_j = \bar{V} . \quad (273)$$

But whatever level i of effort is involved, the so-called "motivational force" is the average valence of the performance levels and therefore is independent of effort level. Thus, everything goes down the drain at once. It is difficult to see how, or even why, anyone would suggest that postulate 3 is in any way relevant to expectancy theory models or how the validity of the formulations could or, much less, would have to depend on it.

The authors conclude their discussion of this postulate with, "If, as seems likely, idiosyncratic, curvilinear relationships are found (between expectancy and valence), research should be directed toward identifying the nature of the relationships and the factors leading to specific types of relationships." We suggest that such efforts could be useful only if they could lead to a modification of the mathematical equations expressing the criterion behavior variables as a function of expectancies and valences of levels of dimensions of specified predictor modalities. In general, a mathematical model can be made sufficiently general with respect to variables, parameters, loss function, and constraints to identify "idiosyncratic curvilinear relationships" to the extent that they are present in the data. The authors, however, like many psychologists, do not specify clearly the structural elements of the system they are considering. Are they considering merely one dimension of the person modality? If so, are they interested primarily in discovering his idiosyncratic curvilinear relationships? Then how would they generalize to other dimensions of the person modality? Or would they attempt to identify the expectancy-valence function for each of a number of dimensions? Having done so, would they have

strategies for studying the individual differences among the functions or for generalizing from them?

Even in the case of the formula, $r_{E_{ij}, V_j} = 0$, do the authors and their colleagues assert that the independence postulate applies to only a single dimension of the person modality across dimensions of an outcome modality or across levels of a single outcome modality? Would they perhaps also entertain the application of the postulate to a single dimension or level of a dimension of an outcome modality but across dimensions of the person modality?

These questions are by no means contentious or frivolous, for they point up the futility of attempting an analysis of any multivariate system without first specifying the elements of both the structural and functional components of the system. It does not appear that postulate 3 is necessarily relevant to expectancy theory. We go then to the next postulate.

Dominance

Postulate 4. "Dominance. If V_1 is greater than V_2 , then the certainty of 1 ($E_1 = 1.00$) has greater motivational force than any combination where $E_1 + E_2 = 1.00$."

If we let F be motivational force, then in the traditional model we have

$$F = E_1 V_1 + E_2 V_2 . \quad (274)$$

According to the postulate, we have

$$V_1 > V_2 \quad (275)$$

$$E_1 + E_2 = 1 . \quad (276)$$

From (274) and (276),

$$F = E_1 (V_1 - V_2) + V_2 . \quad (277)$$

Suppose now we let

$$d = V_1 - V_2 . \quad (278)$$

Because of (275)

$$d > 0 . \quad (279)$$

Consider two different values of E_1 , say E_a and E_b , where

$$E_a - E_b = e \quad (280)$$

and

$$e > 0 . \quad (281)$$

Let F_a and F_b correspond to E_a and E_b respectively. Then we can write, from (277) and (278),

$$F_a = E_a d + V_2 \quad (282)$$

$$F_b = E_b d + V_2. \quad (283)$$

From (280), (282), and (283),

$$F_a - F_b = e d. \quad (284)$$

According to (284), if (275) and (276) hold, a necessary and sufficient condition that $F_a > F_b$ is given by (280) and (281). A special case is when E_a is 1 as in the postulate, for then E_b must be smaller. Hence the postulate is a special case of an algebraic tautology.

The authors follow the statement of the postulate with, "Intrinsic outcomes cannot be ignored in any comprehensive treatment of work motivation; yet the expectancies and the valences associated with them cannot be defined independently of one another. Valence cannot be specified without first specifying the value of expectancy."

The first part of the first statement is obviously true. The second part of the first statement and the second statement are far from obvious. To illustrate their point, the authors give the following example:

Would you prefer a situation in which you would receive

1. A promotion to sales manager if sales in your district exceed \$1 million dollars next year.
2. A promotion to sales manager if sales in your district exceed \$2 million.

Here we have two levels of performance, each leading to the same outcome. Presumably, by definition, the expectancy that each performance level will lead to the outcome (promotion) is equal and equal to the objective probability of 1.00. But what is the expectancy of achieving either performance level and what are the intrinsic valences of these levels? The example cited is silent on these points.

In any case, we find it difficult to relate the authors' discussion of the postulate to the algebraic tautology of which it is a special case. We suggest that this postulate is not relevant to expectancy theory; therefore, we proceed to the next one.

Optimization

Postulate 5. "Optimization. Individuals optimize $\Sigma (E_{ij} V_{ij})$, that is, they cognitively or acognitively compare $\Sigma (E_{ij} V_{ij})_1$ with $\Sigma (E_{ij} V_{ij})_2, \dots, \Sigma (E_{ij} V_{ij})_n$ and pick the best one rather than sequentially comparing each one to some absolute standard derived from their level of aspiration and either (a) ceasing search on identification of an acceptable alternative or (b) being indifferent among those that exceed the standard of satisfactoriness."

Again, as in the case of postulates 1 and 2, we must assume that the subscripts on the closing parentheses refer to the choice or alternative acts i . If so, then we must also assume that it is either the valences or estimated valences of the acts i that are being chosen or compared and on the basis of which an act is chosen. Although the difference between reported and estimated valences has not been recognized in the literature, the issue is important and has been discussed, although far from fully, elsewhere in this document. But here we are concerned only with the choice of one among a number of alternatives.

The question appears to be whether a person continues to search for his most highly valent alternative or searches only until he finds an adequate one. This question suggests that the searching and comparing activity itself might be conceptualized as a distinct dimension of the effort modality whose various levels have different valences. High search effort may have a negative valence and thus reduce the search time in order to optimize the resultant valence of the chosen alternative.

Such an interpretation would bring the satisficing concept into an expectancy theory model with a multidimensional effort modality. There appears to be no indication in the authors' discussion of this postulate that would indicate serious disagreement with this point of view. If the postulate were revised to include "searching" as an additional dimension of effort, it might well be both acceptable and valid. As in all cases involving behavior of human organisms, one would expect to find great individual differences with respect to optimization vs. satisficing.

Reconstitution

Postulate 6. "Reconstitution. Compound estimates can be decomposed and reconstituted according to the rules of probability without affecting $\Sigma (E_{ij} V_j)$. Specifically,

- a. $\Sigma (E_{ij} V_j)$ is unchanged if E_{e+p} and E_{p+r} are exchanged.
- b. $\Sigma (E_{ij} V_j)$ is unchanged if E_1 and E_2 are the expectancies associated with any extrinsic or intrinsic outcome."

First let us consider part (a) of the postulate. Since the authors do not include an adequate definition of notation, we must introduce our own. We shall overlook the failure of the authors to recognize the need for at least two levels for each of the three dimensions (effort, performance, and reward) and assume that consideration is restricted to a single unidentified level of each of the three. We let

- E_{ep} be the expectation originally assigned that the specified level of effort will be followed by the specified level of performance;
- E_{pr} be the expectation originally assigned that the specified level of performance will be followed by the specified level of reward;
- V_p be the valence on the specified level of performance;
- V_r be the specified level of reward;
- F_1 be that component of the force to perform involving only the single levels from each of the three dimensions as calculated from the originally assigned values of E_{ep} and E_{pr} ;

F_2 be defined as F_1 except that the values assigned to E_{ep} and E_{pr} are reversed.

Then, according to the definitions,

$$F_1 = E_{ep} V_p + E_{pr} V_r \quad (285)$$

$$F_2 = E_{pr} V_p + E_{ep} V_r . \quad (286)$$

According to part (a) of the postulate,

$$F_1 = F_2 . \quad (287)$$

From (285), (286), and (287),

$$(E_{ep} - E_{pr}) V_p = (E_{pr} - E_{ep}) V_r . \quad (288)$$

From (288),

$$(E_{ep} - E_{pr}) (V_p + V_r) = 0 . \quad (289)$$

But (289) can hold only if the two expectancies are equal or the two valences are equal in absolute value and opposite in sign. Therefore, it would appear that part (a) of the postulate is mathematically impossible except for the trivial cases just mentioned.

Part (b) of the postulate can fare no better than part (a). Again, we define our notation with the aid of the authors' textual explanation. We let

E_{e1} be the originally assigned expectancy that effort will be followed by outcome 1

E_{e2} be the originally assigned expectancy that effort will be followed by outcome 2

V_1 be the valence of outcome 1

V_2 be the valence of outcome 2.

By analogy, with part (a), we get the part (b) result that

$$(E_{e1} - E_{e2}) (V_1 + V_2) = 0 . \quad (290)$$

Here again, we see that, only in the trivial case of equal expectancies or equal absolute and opposite sign valences can equation (290) hold. We must therefore reject postulate 6 and go on to postulate 7.

Equivalent n-Fold Partition

Postulate 7. "Equivalent n-Fold Partition. Individuals can and do partition the dimension on which they cognitively or acognitively evaluate $\Sigma (E_{ij} V_j)$'s into n or more divisions where n is the number of different levels of work effort they can differentiate."

We suggest that the product summation is meaningless except as a number calculated from numerical values assigned to the E_{ij} 's and V_j 's. For both expectancy and valence dimensions, a number of levels may be specified and values assigned to them. From these, an assignor may assign values for an assignee (the two may be the same). In practice, such a procedure may well yield a wide range of product summation values. Such results may be compared with ratings of effort levels directly reported. We suggest that this postulate reflects a fundamental misconceptualization of any viable expectancy-valence model. We therefore consider next postulate 8.

Irrelevance of Identical Outcomes

Postulate 8. "Irrelevance of Identical Outcomes. Any outcome which is equally likely for all levels of work effort has no impact on the relative preferences among several $\Sigma (E_{ij} V_j)$'s."

Again we note the failure to recognize $\Sigma (E_{ij} V_j)$ as a purely computational expression. If there is only one outcome for whatever the level of effort and its expectancy for all levels is the same and the valence of the outcome is constant for all levels, then the expected valences of all levels would be equal. These assumptions, however, ignore the possibility of the valence of the outcome being a function of the valence of the effort level as well as the intrinsic valences of different effort levels. These concepts have not been adequately integrated into expectancy theory models and the postulate renders a service by calling attention to the need for such an integration. Now we come to the final postulate.

Ambiguity

Postulate 9. "Ambiguity. Where individuals are incapable, because of insufficient information, of judging between two alternatives, they assign equal values to them. Specifically,

a. When an individual lacks information that will enable him to judge whether outcome 1 is preferable to outcome 2, $V_1 = V_2$.

b. When an individual lacks information that will enable him to judge whether the probability of outcome 1 is greater than, equal to, or less than the probability of outcome 2, $E_1 = E_2$."

Again, as in postulate 8, the authors have called attention to a variable that could well be important for expectancy theory but which has been neglected in current formulations. Their statement of the postulate does not explicitly recognize "information," "certainty," "confidence," or what not, as a dimension that would normally have a number of levels for both valence and expectancy. It could well be generalized to encompass the level concept. Earlier in our discussion of the expectancy concept, we have briefly considered the "certainty" concept. Obviously, it should also be extended to the valence modality.

THE RATIONALITY ASSUMPTIONS OF WAHBA AND HOUSE

Another paper, because of its fundamental theoretical orientation, also merits some detailed examination. This is an article by Wahba and House (1974), entitled "Expectancy Theory in Work and Motivation: Some Logical and Methodological Issues." According to

the authors, the purposes of the paper were (1) to review the development of expectancy theory as applied to work and motivation, (2) to specify some of the implicit rationality assumptions underlying the theory, and (3) to raise and clarify some of the methodological issues connected with the major constructs of the theory. In the following paragraphs, we shall consider the authors' second purpose.

Comparison of Rationality Assumptions with the Postulates

The rationality assumptions underlying expectancy theory as discussed by the authors overlap the postulates discussed by Behling and Starke. Presumably, House and Wahba use the term "assumption" in essentially the same way as Behling and Starke use "postulates." The treatment of "rationality" is enlightening because it is defined in terms of clearly stated assumptions. One occasionally encounters criticisms of expectancy theory because the critics question the validity of the rationality assumptions. We suggest that an expectancy theory model need not stand or fall on the extent to which it is based on the stated rationality assumptions. In fact, our discussion of the Behling and Starke expectancy theory postulates indicates our position that there is no logical necessity why they should all hold exactly. We shall then take up the assumptions in the order presented by Behling and Starke:

1. Preference or indifference between alternatives. This assumption is essentially the same as the B-S postulate 1--Comparability.
2. Transitivity of preferences and indifferences. This is essentially the same as the B-S postulate 2--Transitivity.
3. Dominance. This is the same as the B-S postulate 4.
4. Independence of irrelevant outcomes. The assumption appears to be essentially the same as the B-S postulate 8--Irrelevance of identical outcomes.
5. Continuity. Since this assumption does not appear to be represented in Behling and Starke's postulates, it will be discussed in a following section.
6. The final rationality assumption discussed by the authors is "Independence of expectancy and valence." This assumption appears to be the same as B-S postulate 3--Independence.

Continuity Assumption

Assumption 5, Continuity, is not so readily identified with any of the Behling and Starke postulates. We shall therefore quote it verbatim and consider it in some detail.

The Statement of the Assumption

Let outcomes O_1 and O_3 result from action A_1 . Let outcome O_2 result from action B. Finally, let O_2 be preferred to O_1 and O_3 be preferred to O_2 . This assumption states that, under the above conditions, there is a specific set of probabilities that can be assigned to outcomes O_1 and O_3 such that the decision maker will have no preference for action A or B if action B will result in O_2 for certain.

As it stands, this statement of the assumption is not entirely clear and unambiguous. Recognizing this, the authors present an example that we also quote verbatim:

For example, assume that an employee has a choice between (a) working on a number of sales on which he is paid by straight sales commission and can earn either \$500 (O_1) or zero dollars (O_3) per week, depending on how many sales he successfully completes, or (b) a straight salaried assignment on which he is assured of \$150 (O_2) during the same period. The assumption of continuity would predict that there is a specific set of probabilities that can be assigned to the outcome O_1 (\$500 earnings) and outcome O_3 (zero dollars) such that he will have no preference between actions A and B.

If this assumption does not hold empirically, i.e., if there is no such set of expectancies that can be assigned to O_1 and O_3 , then there is no empirical justification for the assumed relationship between valence and expectancy as proposed by the theory.

As in most of the literature on expectancy theory, this statement and its accompanying example do not clearly distinguish the concepts of modalities, dimensions, and levels. It affords a particularly good opportunity to demonstrate the importance of clear and unambiguous terminology and of the need for specifying the elemental and functional components of multivariate systems.

Critique of the Statement

First, let us examine the outcome modality. The outcomes are designated as O_1 , O_2 , and O_3 . We cannot be sure until we have read the example which of these are dimensions and which are levels. The example indicates that they are all levels of a single dimension, money reward, of an outcome modality. In the example, three levels, O_1 , O_2 , O_3 , have been specified and the number assigned to them in dollars per week are 500, 150, and 0, respectively. It is assumed that O_2 is preferred to O_3 and O_1 to O_2 . Presumably, preferences imply valences. We shall let the valences of the three levels be V_1 , V_2 , V_3 respectively and assume that the relationship between values of valence levels and of outcome levels is monotonic. Then the assumption that $O_1 > O_2 > O_3$ is satisfied.

Next, we take up the modality prior to the money dimension of the outcome modality. From the example, we deduce that this antecedent modality is the unidimensional performance modality, sales. The implication of the example is that the performance modality has only two levels. Level 2 is greater than some specified value X and level 1 is X or less.

But now we are confronted with a concept that we have only briefly considered up to now. This is the objective probability as distinguished from subjective probability, otherwise known as expectancy. Although it is not explicitly pointed out by the authors, the objective probability is 1.0 that the higher level, level 2, of sales performance will be followed by level 1 of the dollar outcome. This is also presumably the expectancy of the salesman that level 2 of performance will be followed by level 1 of dollar outcome. Similarly, both the objective probability and the expectancy are also 1.0 that the lower

level, level 1, of performance will be followed by level 3 for the dollar outcome. Clearly then, the performance-expectancy-outcome paradigm does not apply directly in this case.

So far, we have considered only one part of the total situation; namely, act A. According to the scenario for act B, both the expectancy and the objective probability are 1.0 that either performance level 1 or performance level 2 will be followed by level 2 of the salary outcome (\$150). It becomes increasingly clear, therefore, that, for neither act A nor act B, do the expectancies relating performance levels with outcome levels play a role in predicting behavior in the traditional expectancy theory sense, since they are fixed at either 0 or 1.0 for each combination of performance and outcome level. We must therefore examine act B more closely.

According to the example, even if the sales performance were zero, the salesman would still receive salary outcome 2 (\$150) in act B. Unrealistic as this situation is, it cannot be ruled out unless the model is further specified. As this example stands, therefore, we really have 3 levels of performance rather than only 2. These are now level 1, no sales; level 2, some sales but not more than X; and level 3, more than X sales. Presumably, for the model to make sense, another performance value Y must be set for act B below which no salary outcome (zero dollars) would follow.

Before leaving the continuity assumption, we must examine more closely the meaning of "act A" and "act B" and the "preference for" one or the other. Lacking an explicit definition by the authors of "preference for," we shall assume that, in some way, valence levels have been associated with both acts A and B. However A and B are defined, we assume that the one with the highest valence is preferred. The problem, then, is to state exactly what acts A and B are and how their valences might be estimated.

We begin by assuming that A and B refer to total situations that can be discussed as follows.

Mathematical Formulations of the Assumption

For a given person let

- n_i be the number of levels of a prior dimension
- n_j be the number of levels of a subsequent dimension
- P_i be the expectation level that he can achieve level i of a prior modality dimension
- V_j be the valence level of a subsequent outcome dimension
- $A^{P_{ij}}$ be the expectation level that level i of a prior modality dimension will be followed by level j of a subsequent modality dimension for situation A
- $B^{P_{ij}}$ be the same as $A^{P_{ij}}$ for situation B
- P be the n_i th order vector of P_i 's
- V be the n_j th order vector of V_j 's
- A^P be the $n_i \times n_j$ matrix of $A^{P_{ij}}$

${}_B P$ be the $n_i \times n_j$ matrix of ${}_B P_{ij}$

\tilde{V}_A be the estimated valence for situation A

\tilde{V}_B be the estimated valence for situation B.

Then the expected valences \tilde{V}_A and \tilde{V}_B , respectively, are:

$$\tilde{V}_A = P' {}_A P V \quad (291)$$

$$\tilde{V}_B = P' {}_B P V. \quad (292)$$

We assume that ${}_A P$ and ${}_B P$ are the objective as well as the subjective probabilities. Presumably, they are fixed by management and understood by the subject. We also assume that $V_{j+1} > V_j$. This merely means that a higher level outcome has a higher valence than a lower level outcome. The continuity assumption, then, appears to be that the ${}_A P$ and ${}_B P$ matrices and the P vector can be determined in such a way that

$$\tilde{V}_A = \tilde{V}_B. \quad (293)$$

For the example given, we have

$${}_A P \equiv \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad (294)$$

$${}_B P \equiv \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad (295)$$

We let

$$P \equiv \begin{bmatrix} P_1 \\ P_0 \end{bmatrix} \quad (296)$$

$$V \equiv \begin{bmatrix} V_3 \\ V_2 \\ V_1 \end{bmatrix}. \quad (297)$$

By definition

$$P_1 + P_0 = 1. \quad (298)$$

From equation (291) through (298), we get

$$\tilde{V}_A = P_1 (V_3 - V_0) + V_0 \quad (299)$$

$$\tilde{V}_B = V_2. \quad (300)$$

The assumption, then, is that the sales levels corresponding to P_1 and P_0 can be determined so that (299) and (300) yield the same expected valences \tilde{V}_A and \tilde{V}_B . It seems highly unlikely that for most persons cutting points could not be established to yield a P_1 satisfying $\tilde{V}_A = \tilde{V}_B$.

From a theoretical point of view then, the continuity assumption, as described by Wahba and House, would appear to be easily satisfied and does not seem to be a significant issue in expectancy theory. From a practical point of view, it may have considerable importance in cost-effective considerations. The determination of the objective conditional probability matrix and the expectancy performance vector so as to optimize some specified criterion objective could be of great practical interest.

Weaknesses of the Continuity Statement

Before leaving the continuity assumption, we note briefly several weaknesses in the authors' statement of the assumption.

1. They fail to recognize explicitly the distinction between objective and subjective probabilities.
2. They fail to specify clearly the distinction between dimensions and levels of a dimension of an outcome modality.
3. They fail to recognize explicitly that each dimension of a modality must have at least two levels.
4. They do not distinguish between the levels of a dimension, the values assigned to the levels, and the valence levels assigned to the dimension levels.
5. They do not distinguish between joint probabilities and conditional probabilities.
6. They confuse the objective conditional probability of a level of a dimension of a subsequent modality with the expectation of a level of a dimension of a prior modality.
7. They do not distinguish clearly between the value of a level of a dimension of a prior modality and the expectancy level of that dimension level.
8. They indicate no recognition of the important fact that the probability matrices implied in their example are treated as intervention or manipulable variables.

Although there may be other significant oversights in the authors' statement and discussion of the continuity assumption, the above eight are doubtless among the most relevant for a comprehensive, well-structured expectancy theory model.

Except for the continuity assumption, the Wahba and House rationality assumptions are essentially included in the Behling and Starke postulates. Therefore, our discussion of the latter may be taken as applying to the former as well.

In any case, our discussion of these postulates and assumptions is not intended as a criticism of the important contributions of these thoughtful researchers but, rather, as a means of emphasizing the need for a more adequately formulated expectancy theory model.

CONCLUSIONS

It must be evident that this document has made no more than meager reference to the vast amount of experimental and applied research conducted in the area over the past 20 years. Although no pretense is made of familiarity with all the details of this impressive effort, a respectable sampling of the work has been consulted. This has led to the conclusion that most, if not all, of the work could have profited substantially from modeling based on better taxonomic structuring and more rigorous mathematical formulations.

In our opinion, this statement is equally applicable to a great deal of past and current psychological research, as well as research and application in the other social sciences. However, expectancy theory offers a particularly good challenge to apply the best available resources for taxonomic and mathematical modeling. This is true because it addresses the highly important area of motivation that has been well supplied with literary, semantic, and rhetorical effort but supported at best only modestly by a rigorous taxonomic and mathematical infrastructure.

Certainly, the mathematical models suggested or proposed may well be improved. Also, the current state of the art has certainly demonstrated the practical value of expectancy theory in organizational behavior. It is not suggested that further applied work be halted until more theoretically elegant or adequate models have been fully developed and tested. However, it is believed that, along with applied efforts, it is important that this development and testing of more scientifically rigorous models should proceed.

It is believed that an important thrust for modeling research should be directed toward simplification of the responses required from persons to controlled format inventories. One principle is to let the model solve for scale and origin parameters and relieve the subject of making these judgments. In expectancy theory experiments involving absolute judgments with respect to probabilities and valences, it is particularly tempting to impose on the subject the burden of making absolute judgments of origin and scale. In this and other ways, expectancy theory modeling can offer opportunities for developing designs in which the model minimizes the effort of the subject. The paired matrix models given in Appendices A through H illustrate approaches that can contribute toward this goal. In general, in research or evaluation projects involving human subjects, the effort required by the subject should be minimized unless this effort is an essential aspect to be investigated by the project.

Appendices A through H were all developed for designs yielding two separate matrices of measures that purport to yield the same results, except for origin and scaling of the subjects' responses. The rows represent persons and the columns represent activities. The various models represent variations in constraints imposed for the scaling and origin parameter vectors. Other experimental designs may well be modeled. Some or all of these should be applied to experimental data.

REFERENCES

- Behling, O. & Starke, F. A. The postulates of expectancy theory. Academy of Management Journal, 16, 373-388, 1973.
- Burket, G. R. A study of reduced rank models for multiple prediction. Psychometric Monographs, No. 12, 1964.
- Campbell, J. P., Dunnette, M. D., Lawler, E. E., III, & Welch, K. E., Jr. Managerial behavior, performance, and effectiveness. New York: McGraw-Hill, 1970.
- Campbell, J. P., & Pritchard, R. D. Motivation theory in industrial and organizational psychology. In M. D. Dunnette, ed., Handbook of industrial and organization psychology. Chicago: Rand McNally, 63-130, 1976.
- Campbell, M. M. The primary dimensions of item ratings on scales designed to measure twenty-four of Murray's Manifest Needs. Unpublished doctoral dissertation, University of Washington, 1959.
- Cattell, R. B. The data box: Its ordering of total resources in terms of possible relational systems. In Handbook of multivariate experimental psychology. Chicago: Rand McNally, 67-128, 1966.
- Coombs, C. H. A theory of data. New York: Wiley, 1964.
- Dashiell, J. F. Fundamentals of general psychology. Boston: Houghton Mifflin, 1949.
- Dockstader, S. L., Nebeker, D. M., Nocella, J., & Shumate, E. C. Incentive management training: Use of behavioral principles for productivity enhancement (NPRDC Tech. Rep. 80-29). San Diego: Navy Personnel Research and Development Center, July 1980. (AD-A087 489).
- Edwards, A. L. Edwards personal preference schedule. New York: Psychological Corporation, 1953.
- Galbraith, J., & Cummings, L. L. An empirical investigation of the motivational determinants of task performance: Interactive effects between instrumentality-valence and motivation-ability. Organizational Behavior and Human Performance, 2, 237-257, 1967.
- Gulliksen, H. A rational equation of the learning curve based on Thorndike's law of effect. Journal of General Psychology, 11, 395-434, 1934.
- Guttman, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 33, No. 4, 469-506, 1968.
- Horst, P. A method for determining the absolute affective value of a series of stimulus situations. Journal of Educational Psychology, Vol. XXIII, No. 6, September 1932.
- Horst, P. Multivariate models for evaluating change. In Chester Harris, et al., Problems in measuring change. Madison, WI: University of Wisconsin Press, 104-121, 1963.
- Horst, P. Factor analysis of data matrices. New York: Holt, Rinehart, & Winston, 1965.

- Horst, P. Psychological measurement and prediction. Belmont, CA: Wadsworth Publishing Company, 1966.
- Horst, P. Reduced rank regression analysis for optimal cross validity (University of Washington Rep.), September 1970.
- Horst, P. Prediction from multivariate polynomial variables. ORI Technical Report, 13(1), 1973.
- Hotelling, H. Relations between two sets of variates. Biometrika, 28, 321-377, 1936.
- Ilgen, D. R., Pritchard, R. D., Dugoni, B. L., Matte, Jr., W. H., & Nebeker, D. M. An empirical study of the effectiveness of several expectancy theory measures in an experimental work simulation (NPRDC Tech. Note 80-20). San Diego: Navy Personnel Research and Development Center, July 1980.
- Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29, 115-129, 1964.
- Lazarusfeld, P. F. The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer, et al., Measurement and prediction. Princeton: Princeton University Press, Chapter 10, 1950.
- Leiman, J. The calculation of regression weights from common factor loadings. Unpublished doctoral dissertation, University of Washington, 1951.
- Lingoes, J. C. An IBM 7090 program for Guttman-Lingoes smallest space analysis-I. Behavioral Science, 10, 183-184, 1965.
- Mitchell, T. R. Expectancy models of job satisfaction, occupational preference and effort. Psychological Bulletin, 81(12), 1053-1077, 1974.
- Murray, H. A., et al., Explorations in personality. New York: Oxford University Press, 1938.
- Nebeker, D. M., & Mitchell, T. R. Alternative interpretations of measurement problems associated with expectancy theory: A response to Schmidt (Unpublished Working Paper). Seattle, WA: University of Washington, 1973.
- Nebeker, D. M., & Moy, M. C. Work performance: A new approach to expectancy theory predictions (NPRDC Tech. Rep. 76TQ-47). San Diego: Navy Personnel Research and Development Center, September 1976. (AD-A030 451).
- Schmidt, F. L. Implications of a measurement problem for expectancy theory research. Organizational Behavior and Human Performance, 10, 243-251, 1973.
- Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I, II. Psychometrika, 27, 125-140, 219-246, 1962.
- Thurstone, L. L. The learning function. Journal of General Psychology, 3, 469-493, 1930.
- Thurstone, L. L. A law of comparative judgment. Psychological Review, 34, 273-286, 1927.

Tucker, L. R. Implications of factor analysis of three-way matrices for measurement of change. In Chester Harris, et al., Problems in Measuring Change. Madison, WI: University of Wisconsin Press, 122-137, 1963.

Vroom, V. H. Work and motivation. New York: Wiley, 1964.

Wahba, M. A., & House R. J. Expectancy theory in work and motivation: Some logical and methodological issues. Human Relations, 27, 121-147, 1974.

APPENDIX A
OPTIMAL ROW TRANSFORMATIONS OF PAIRED MATRICES

APPENDIX A

OPTIMAL ROW TRANSFORMATIONS OF PAIRED MATRICES

I. The General Model

Suppose we have given two $n \times m$ matrices of measures X and Y which purport to measure the same functions except that, for some or all of the rows of each matrix, unknown linear transformations may have taken place. The additive and multiplying constants of these row transformations may be regarded as characteristics of the row entities and are to be solved for. We wish to determine them so that for the transformed matrices, say x and y , the sum of the correlations of corresponding columns of the two matrices will be a maximum. If the matrices do indeed measure the same functions except for linear row transformations, then the correlations between corresponding rows of the two matrices should be unity for all rows. In any case, correlations between corresponding rows of X and Y will be independent of any linear transformations of the rows of either or both matrices. We shall therefore consider the determination of row transformation parameters which will maximize the average of corresponding column correlations of the two matrices.

We let

X be an $n \times m$ matrix of experimentally determined measures

Y be an $n \times m$ matrix of experimentally determined measures

D_X be an n th order diagonal matrix of scaling constants for rows of X

D_Y be an n th order diagonal matrix of scaling constants for rows of Y

B_X be an n th order vector of additive constants for rows of X

B_Y be an n th order vector of additive constants for rows of Y

$$x = D_X X + B_X 1' \tag{1.1}$$

$$y = D_Y Y + B_Y 1' \tag{1.2}$$

We see then that the x and y matrices in equations 1.1 and 1.2, respectively, have been obtained by linear row transformations on the X and Y matrices, respectively.

We let

$x_{.j}$ be the j th column vector of x

$y_{.j}$ be the j th column vector of y

$X_{.j}$ and $Y_{.j}$ be similarly defined

r_j be the correlation between columns j of x and y

$$A_X = D_X 1 \quad (1.3)$$

$$A_Y = D_Y 1 \quad (1.4)$$

Now from equations 1.1 and 1.2

$$x_{.j} = D_X X_{.j} + B_X \quad (1.5)$$

$$y_{.j} = D_Y Y_{.j} + B_Y \quad (1.6)$$

If we let

$D_{X_{.j}}$ be a diagonal matrix of $X_{.j}$

$D_{Y_{.j}}$ be a diagonal matrix of $Y_{.j}$

we may, because of equations 1.3 and 1.4, write equations 1.5 and 1.6, respectively,

$$x_{.j} = (D_{X_{.j}}, I) \begin{bmatrix} A_X \\ B_X \end{bmatrix} \quad (1.7)$$

$$y_{.j} = (D_{Y_{.j}}, I) \begin{bmatrix} A_Y \\ B_Y \end{bmatrix} \quad (1.8)$$

Suppose we let

c_j be the covariance between $x_{.j}$ and $y_{.j}$

v_{x_j} be the variance of $x_{.j}$

v_{y_j} be the variance of $y_{.j}$

Then from the definitions

$$c_j = X_{.j}' (I - \frac{1 1'}{n}) Y_{.j} / n \quad (1.9)$$

$$v_{x_j} = X_{.j}' (I - \frac{1 1'}{n}) X_{.j} / n \quad (1.10)$$

$$v_{y_j} = Y_{.j}' (I - \frac{1 \cdot 1'}{n}) Y_{.j} / n \quad (1.11)$$

Also, from the definitions

$$r_j = \frac{c_j}{v_{x_j}^{\frac{1}{2}} v_{y_j}^{\frac{1}{2}}} \quad (1.12)$$

We let

$$\theta = \sum r_j / m \quad (1.13)$$

II. The General Solution

We wish now to determine A_X , B_X and A_Y , B_Y in equations 1.7 and 1.8, respectively, so that θ in equation 1.13 is a maximum.

Taking differentials of both sides of equation 1.13, we have

$$d\theta = \sum d r_j / m \quad (2.1)$$

But from equation 1.12

$$d r_j = \frac{d c_j}{v_{x_j}^{\frac{1}{2}} v_{y_j}^{\frac{1}{2}}} - \frac{1}{2} \frac{c_j}{v_{x_j}^{\frac{1}{2}} v_{y_j}^{\frac{1}{2}}} \left[\frac{d v_{x_j}}{v_{x_j}} + \frac{d v_{y_j}}{v_{y_j}} \right] \quad (2.2)$$

Suppose now we define

$$M = (I - \frac{1 \cdot 1'}{n}) \quad (2.3)$$

$$\mu_X' = (A_X', B_X') \quad (2.4)$$

$$\mu_Y' = (A_Y', B_Y') \quad (2.5)$$

$$v_{X_j}' = (D_{X_{.j}}, I) \quad (2.6)$$

$$v_{Y_j}' = (D_{Y_{.j}}, I) \quad (2.7)$$

From equations 1.7 through 1.12 and 2.3 through 2.7

$$c_j = \mu_X' v_{X_j} M v_{Y_j}' \mu_Y' / n \quad (2.8)$$

$$v_{x_j} = \mu_X' v_{X_j} M v_{X_j}' \mu_X' / n \quad (2.9)$$

$$v_{y_j} = \mu_Y' v_{Y_j} M v_{Y_j}' \mu_Y' / n \quad (2.10)$$

Next define

$$j_{XY}^d = \frac{1}{(v_{xj} v_{yj})^{\frac{1}{2}}} \quad (2.11)$$

$$j_X^d = r_j / v_{xj} \quad (2.12)$$

$$j_Y^d = r_j / v_{yj} \quad (2.13)$$

and

$$d_{XY} = \begin{bmatrix} j_{XY}^d & \cdot & \cdot & 0 \\ 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & j_{XY}^d \end{bmatrix} \quad (2.14)$$

$$d_X = \begin{bmatrix} j_X^d & \cdot & \cdot & 0 \\ 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & j_X^d \end{bmatrix} \quad (2.15)$$

$$d_Y = \begin{bmatrix} j_Y^d & \cdot & \cdot & 0 \\ 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & j_Y^d \end{bmatrix} \quad (2.16)$$

$$v_{XY} = d_{XY}^{-1} \quad (2.17)$$

$$v_X = d_X^{-1} \quad (2.18)$$

$$v_Y = d_Y^{-1} \quad (2.19)$$

Now to maximize θ in equation 1.13, we write

$$\frac{\partial \theta}{\partial \mu_X} = 0 \quad (2.20)$$

$$\frac{\partial \theta}{\partial \mu_Y} = 0 \quad (2.21)$$

But from equation 2.1

$$\frac{\partial \theta}{\partial \mu_X} = \sum \frac{\partial r_j}{\partial \mu_X} \quad (2.22)$$

$$\frac{\partial \theta}{\partial \mu_Y} = \sum \frac{\partial r_j}{\partial \mu_Y} \quad (2.23)$$

From equations 2.2, and 2.11 through 2.13

$$\frac{\partial r_j}{\partial \mu_X} = j_{XY}^d \frac{\partial c_j}{\partial \mu_X} - \frac{1}{2} \left(j_X^d \frac{\partial v_{xj}}{\partial \mu_X} + j_Y^d \frac{\partial v_{yj}}{\partial \mu_X} \right) \quad (2.24)$$

$$\frac{\partial r_j}{\partial \mu_Y'} = j_{XY} \frac{\partial c_j}{\partial \mu_Y'} - \frac{1}{2} (j_{Xj} \frac{\partial v_{Xj}}{\partial \mu_Y'} + j_{Yj} \frac{\partial v_{Yj}}{\partial \mu_Y'}) \quad (2.25)$$

From equation 2.8

$$\frac{\partial c_j}{\partial \mu_X'} = v_{Xj} \text{ M } v_{Yj}' \mu_Y / n \quad (2.26)$$

$$\frac{\partial c_j}{\partial \mu_Y'} = v_{Yj} \text{ M } v_{Xj}' \mu_X / n \quad (2.27)$$

From equation 2.9

$$\frac{\partial v_{Xj}}{\partial \mu_X'} = 2 v_{Xj} \text{ M } v_{Xj}' \mu_X / n \quad (2.28)$$

$$\frac{\partial v_{Xj}}{\partial \mu_Y'} = 0 \quad (2.29)$$

From equation 2.10

$$\frac{\partial v_{Yj}}{\partial \mu_X'} = 0 \quad (2.30)$$

$$\frac{\partial v_{Yj}}{\partial \mu_Y'} = 2 v_{Yj} \text{ M } v_{Yj}' \mu_Y / n \quad (2.31)$$

From equations 2.24, 2.26, 2.28, and 2.30

$$\frac{\partial r_j}{\partial \mu_X'} = j_{XY} v_{Xj} \text{ M } v_{Yj}' \mu_Y / n - j_{Xj} v_{Xj} \text{ M } v_{Xj}' \mu_X / n \quad (2.32)$$

From equations 2.25, 2.27, 2.29, and 2.31

$$\frac{\partial r_j}{\partial \mu_Y'} = j_{XY} v_{Yj} \text{ M } v_{Xj}' \mu_X / n - j_{Yj} v_{Yj} \text{ M } v_{Yj}' \mu_Y / n \quad (2.33)$$

Let

$$j_{XX} = v_{Xj} \text{ M } v_{Xj}' \quad (2.34)$$

$$j_{XY} = v_{Xj} \text{ M } v_{Yj}' \quad (2.35)$$

$$j_{YY} = v_{Yj} \text{ M } v_{Yj}' \quad (2.36)$$

From equations 2.20, 2.22, 2.32, 2.34, and 2.35

$$0 = \sum_j (j^d_X j^{\phi}_{XX} \mu_X - j^d_{XY} j^{\phi}_{XY} \mu_Y) \quad (2.37)$$

$$0 = \sum_j (-j^d_{XY} j^{\phi}_{YX} \mu_X + j^d_Y j^{\phi}_{YY} \mu_Y) \quad (2.38)$$

From equations 2.37 and 2.38 we may write

$$\begin{bmatrix} \sum j^d_X j^{\phi}_{XX} & -\sum j^d_{XY} j^{\phi}_{XY} \\ -\sum j^d_{XY} j^{\phi}_{YX} & \sum j^d_Y j^{\phi}_{YY} \end{bmatrix} \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (2.39)$$

We now consider the solution for the μ 's in equation 2.39. We note that the j^d 's in equation 2.39 are functions of μ_X and μ_Y . Assume that we have some initial approximation to the μ 's and that we calculate approximations to the j^d 's by means of equations 2.8 through 2.10. But the j^d_X 's and j^d_Y 's are functions also of the r_j 's. We assume that the approximated r_j 's are proportional to the true r_j 's by a proportionality constant k . We let

$$\sigma_{XX} = \sum j^d_X j^{\phi}_{XX} \quad (2.40)$$

$$\sigma_{XY} = \sum j^d_{XY} j^{\phi}_{XY} \quad (2.41)$$

$$\sigma_{YY} = \sum j^d_Y j^{\phi}_{YY} \quad (2.42)$$

and write from equations 2.39 through 2.42

$$\begin{bmatrix} k \sigma_{XX} & -\sigma_{XY} \\ -\sigma_{YX} & k \sigma_{YY} \end{bmatrix} \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (2.43)$$

Now equation 2.43 is the well known form for the canonical correlation between two sets of variables. But suppose we define

$$f' = (0', 1') / \sqrt{m} \quad (2.44)$$

where 0 and 1 are each of order m . From equations 2.3, 2.6, 2.7, 2.34, 2.35, 2.36, 2.40, 2.41, and 2.42, it can be proved that

$$\sigma_{XX} f = 0 \quad (2.45)$$

$$\sigma_{XY} f = 0 \quad (2.46)$$

$$\sigma_{YX} f = 0 \quad (2.47)$$

$$\sigma_{YY} f = 0 \quad (2.48)$$

Therefore, from equations 2.45 through 2.48, all the σ 's are non-basic. They are of order m and, except for special cases, of rank $m - 1$, and hence the conventional solutions for equation 2.43 are not available.

III. The Non-Basic Problem

We shall therefore consider the special case indicated by the conditions in equations 2.45 through 2.48. We write the basic structure forms (see Horst, 1963)*

$$\sigma_{XX} = Q_X \pi_X^2 Q_X' \quad (3.1)$$

$$\sigma_{XY} = q_X D q_Y' \quad (3.2)$$

$$\sigma_{YY} = Q_Y \pi_Y^2 Q_Y' \quad (3.3)$$

Now let

$$H_X \equiv (Q_X, f) \quad (3.4)$$

$$H_Y \equiv (Q_Y, f) \quad (3.5)$$

$$h_X \equiv (h_X, f) \quad (3.6)$$

$$h_Y \equiv (h_Y, f) \quad (3.7)$$

$$\delta_X \equiv \begin{bmatrix} \pi_X & 0 \\ 0 & 0 \end{bmatrix} \quad (3.8)$$

$$\delta_Y \equiv \begin{bmatrix} \pi_Y & 0 \\ 0 & 0 \end{bmatrix} \quad (3.9)$$

$$d \equiv \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \quad (3.10)$$

$$\Sigma_{XX} = H_X \delta_X^2 H_X' \quad (3.11)$$

$$\Sigma_{YY} = H_Y \delta_Y^2 H_Y' \quad (3.12)$$

$$\Sigma_{XY} = h_X D h_Y' \quad (3.13)$$

From equations 2.45 through 3.3 and 3.4 through 3.7, the H 's and h 's are square orthonormal. Also, from equations 3.1 through 3.13

$$\sigma_{XX} = \Sigma_{XX} \quad (3.14)$$

$$\sigma_{XY} = \Sigma_{XY} \quad (3.15)$$

$$\sigma_{YY} = \Sigma_{YY} \quad (3.16)$$

From equations 2.43 and 3.11 through 3.16

$$\begin{bmatrix} k H_X \delta_X^2 H_X' & -h_X D h_Y' \\ -h_Y D h_X' & k H_Y \delta_Y^2 H_Y' \end{bmatrix} \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (3.17)$$

From equation 3.17

$$\begin{bmatrix} k \delta_X^2 & -H_X' h_X D h_Y' H_Y \\ -H_Y' h_Y D h_X' H_X & k \delta_Y^2 \end{bmatrix} \begin{bmatrix} H_X' \mu_X \\ H_Y' \mu_Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (3.18)$$

From equations 2.45 through 3.3

$$H_X' h_X = \begin{bmatrix} q_X' q_X & 0 \\ 0 & 1 \end{bmatrix} \quad (3.19)$$

$$H_Y' h_Y = \begin{bmatrix} q_Y' q_Y & 0 \\ 0 & 1 \end{bmatrix} \quad (3.20)$$

From equations 3.4 through 3.10 and 3.13 through 3.20

$$\begin{bmatrix} k \begin{bmatrix} \pi_X^2 & 0 \\ 0 & 0 \end{bmatrix} & - \begin{bmatrix} q_X' q_X & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} d & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} q_Y' q_Y & 0 \\ 0 & 1 \end{bmatrix} \\ - \begin{bmatrix} q_Y' q_Y & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} d & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} q_X' q_X & 0 \\ 0 & 1 \end{bmatrix} & k \begin{bmatrix} \pi_Y^2 & 0 \\ 0 & 0 \end{bmatrix} \end{bmatrix} \begin{bmatrix} q_X' \mu_X \\ f' \mu_X \\ q_Y' \mu_Y \\ f' \mu_Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (3.21)$$

From equation 3.21

$$f' \mu_X = 0 \quad (3.22)$$

$$f' \mu_Y = 0 \quad (3.23)$$

From equations 3.21, 3.22, and 3.23

$$\begin{bmatrix} k \pi_X^2 & -Q_X' q_X d q_Y' Q_Y \\ -Q_Y' q_Y d q_X' Q_X & k \pi_Y^2 \end{bmatrix} \begin{bmatrix} Q_X' \mu_X \\ Q_Y' \mu_Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (3.24)$$

Let

$$E_{XY} = \pi_X^{-1} Q_X' q_X d q_Y' Q_Y \pi_Y^{-1} \quad (3.25)$$

$$u_X = \pi_X Q_X' \mu_X \quad (3.26)$$

$$u_Y = \pi_Y Q_Y' \mu_Y \quad (3.27)$$

From equations 3.24 through 3.27

$$\begin{bmatrix} k I & -E_{XY} \\ -E_{YX} & k I \end{bmatrix} \begin{bmatrix} u_X \\ u_Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (3.28)$$

From equation 3.28

$$(E_{YX} E_{XY} - k^2 I) u_Y = 0 \quad (3.29)$$

Write the basic structure form

$$E_{YX} = g_Y D_k g_X \quad (3.30)$$

From equations 3.29 and 3.30

$$g_Y D_k^2 q_Y' u_Y = k^2 u_Y \quad (3.31)$$

From equation 3.31 u_Y is a vector of the basic orthonormal of g_Y . Since we wish k to be a maximum, we take the vector of g_Y corresponding to the largest element of D_k^2 for u_Y .

Similarly, we can show that u_X is taken as the corresponding vector of g_X .

To solve for μ_X , we solve for the minimum length vector μ_X which satisfies equation 3.26. This can be shown to be (see Horst, 1963)*

$$\mu_X = Q_X \pi_X^{-1} u_X \quad (3.32)$$

Similarly

$$\mu_Y = Q_Y \pi_Y^{-1} u_Y \quad (3.33)$$

IV. The Iterative Computations

We recall that the foregoing development assumed that we had a first approximation to the μ 's from which we calculated first approximations to the c_j 's, v_{x_j} 's, and v_{y_j} 's of equations 2.8, 2.9, and 2.10. From these we calculated first approximations to the $j^{d_{XY}}$'s, j^{d_X} 's, and j^{d_Y} 's in equations 2.11, 2.12, and 2.13. We then, by means of equations 2.34 through 2.36 and 2.40 through 2.42, calculated first approximations to the σ 's. The basic structure factors * of these σ 's were calculated as indicated in equations 3.1 through 3.3. From these basic structure factors the matrix E_{XY} in equation 3.25 was calculated. The basic orthonormal vectors of E_{XY} corresponding to the largest element of D_k in equation 3.30 were then calculated and indicated by u_X and u_Y , respectively. Finally, from equations 3.32 and 3.33, the μ_X and μ_Y vectors were calculated. These are the second approximations to the μ_X and μ_Y vectors.

Using these second approximations to the μ_X and μ_Y vectors, we may now repeat the foregoing cycle of computations to get a third approximation to the two vectors. These cycles may be continued until the largest element in D_k of equation 3.30 approaches unity, at which time presumably θ in equation 1.13 should be a maximum.

To choose a first approximation to the μ 's, we note first, from equations 2.4, 2.5, 2.44, 3.22, and 3.23, that

$$B_X' 1 = 0 \quad (4.1)$$

$$B_Y' 1 = 0 \quad (4.2)$$

For want of a better estimate, we may also assume that all elements of the B 's are equal. With this assumption we have as a first approximation for the B 's

$$B_X = 0 \quad (4.3)$$

$$B_Y = 0 \quad (4.4)$$

We also note from equations 1.12, 1.13, 2.8, 2.9, and 2.10, that θ is independent of any scaling constant for both μ_X and μ_Y . For want of a better estimate we may assume that all elements of both A_X and A_Y in equation 2.4 and 2.5 are all equal and, for simplicity, that they are all unity. We have therefore as a first approximation to μ_X and μ_Y

$$\mu_X = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (4.5)$$

$$\mu_Y = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (4.6)$$

where the subvectors on the right of equations 4.5 and 4.6 are all of order m .

It must be evident by now that the computations implied by the foregoing development can be excessive for large numbers of entities and variables. We shall therefore not develop in detail a sequence of computational algorithms for the method outlined. Rather, we shall consider an alternative approach that is much simpler in rationale and computational requirements.

V. Computational Simplifications

We begin by assuming, without loss of generality, that only the X matrix has been subjected to linear transformation by rows, and we wish to find the parameters of linear row transformations which will maximize the sum of correlations between corresponding columns of the Y and the row-transformed X matrix. We shall, on occasion, use slightly different notation than in the preceding sections.

We let

$$x_{.j} = (I - \frac{1 \cdot 1'}{n}) (D_{X_{.j}}, I) \begin{bmatrix} A \\ B \end{bmatrix} \quad (5.1)$$

where $x_{.j}$ is the j th column of the row-transformed X matrix, $D_{X_{.j}}$ is a diagonal matrix of the j th column of the original X matrix, A is the vector of proportionality constants for the row transformations of X , and B is the vector of additive constants. It can be seen therefore that $x_{.j}$ in equation 5.1 is a vector of deviated values of the j th column of the transformed X matrix.

We let

$$y_{.j} = (I - \frac{1 \cdot 1'}{n}) Y_{.j} / \sigma_j \quad (5.2)$$

where σ_j is the standard deviation of the j th column of the Y matrix.

Therefore $y_{.j}$ in equation 5.2 is a column of standardized measures for the j th column of Y . We let

$$f_{Y_j} = x_{.j}' y_{.j} / n \quad (5.3)$$

$$f_{X_j} = x_{.j}' x_{.j} / n \quad (5.4)$$

$$r_j = \frac{f_{Y_j}}{f_{X_j}^{1/2}} \quad (5.5)$$

It can readily be seen from the definitions that r_j in equation 5.5 is the correlation between the j th column of the Y matrix and the j th column of the row-transformed X matrix.

We let

$$\theta = \sum_{j=1}^m r_j \quad (5.6)$$

From equation 5.6, θ is the sum of all the r_j correlations. We wish to determine the A and B vectors in equation 5.1 so as to maximize θ in equation 5.6. We begin by taking differentials of both sides of equation 5.6. This gives

$$d\theta = \sum_{j=1}^m d r_j \quad (5.7)$$

From equation 5.5 we get

$$d r_{.j} = \frac{d f_{Y_j}}{f_{X_j}^{1/2}} - \frac{1}{2} \frac{f_{Y_j}}{f_{X_j}^{3/2}} d f_{X_j} \quad (5.8)$$

Taking partial derivatives of θ with respect to A and B , we get from equations 5.7 and 5.8

$$\frac{\partial \theta}{\partial (A', B')} = \sum_{j=1}^m \left(\frac{1}{f_{X_j}^{1/2}} \frac{\partial f_{Y_j}}{\partial (A', B')} - \frac{1}{2} \frac{f_{Y_j}}{f_{X_j}^{3/2}} \frac{\partial f_{X_j}}{\partial (A', B')} \right) \quad (5.9)$$

Now from equations 5.1, 5.2, and 5.3

$$f_{Y_j} = (A', B') \begin{bmatrix} D_{X.j} \\ I \end{bmatrix} (I - \frac{1 \cdot 1'}{n}) y_{.j} / n \quad (5.10)$$

From equations 5.1, 5.2, and 5.4

$$f_{X_j} = (A', B') \begin{bmatrix} D_{X.j} \\ I \end{bmatrix} (I - \frac{1 \cdot 1'}{n}) (D_{X.j}, I) \begin{bmatrix} A \\ B \end{bmatrix} / n \quad (5.11)$$

From equation 5.10

$$\frac{\partial f_{Y_j}}{\partial (A', B')} = \begin{bmatrix} D_{X.j} \\ I \end{bmatrix} (I - \frac{1 \cdot 1'}{n}) y_{.j} / n \quad (5.12)$$

From equation 5.11

$$\frac{\partial f_{X_j}}{\partial (A', B')} = 2 \begin{bmatrix} D_{X.j} \\ I \end{bmatrix} (I - \frac{1 \cdot 1'}{n}) (D_{X.j}, I) \begin{bmatrix} A \\ B \end{bmatrix} / n \quad (5.13)$$

For the maximizing condition, we set

$$\frac{\partial \theta}{\partial (A', B')} = 0 \quad (5.14)$$

From equations 5.9, 5.12, 5.13, and 5.14

$$\sum_{j=1}^n \frac{1}{f_{X_j}^{1/2}} \begin{bmatrix} D_{X.j} \\ I \end{bmatrix} (I - \frac{1 \cdot 1'}{n}) y_{.j} - \sum_{j=1}^n \left[\frac{f_{Y_j}}{f_{X_j}^{3/2}} \begin{bmatrix} D_{X.j} \\ I \end{bmatrix} (I - \frac{1 \cdot 1'}{n}) (D_{X.j}, I) \begin{bmatrix} A \\ B \end{bmatrix} \right] = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (5.15)$$

We now let D_{f_X} be a diagonal matrix of the f_{X_j} 's and D_Y a diagonal matrix of the f_{Y_j} 's and define

$$D_X = D_{f_X}^{-1/2} \quad (5.16)$$

$$D_Y = D_Y D_X^3 \quad (5.17)$$

$$F_X = D_X 1 \quad (5.18)$$

We indicate by $X \cdot y$ the elemental product of the X and y matrices. From equations 5.15, 5.16, 5.17, and 5.18, we have

$$\begin{bmatrix} X \cdot y \\ y \end{bmatrix} F_X = \begin{bmatrix} D_{XD_Y X'} & D_{XD_Y 11'} \\ D_{11' D_Y X'} & D_{11' D_Y 11'} \end{bmatrix} - \begin{bmatrix} X \\ 1 \ 1' \end{bmatrix} D_{Y/n} \begin{bmatrix} X', 1 \ 1' \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (5.19)$$

where the subscripted D 's in the second matrix on the left of equation 5.19 indicate the diagonal of the matrix product subscripts. We let

$$F_Y = D_Y \ 1 \quad (5.20)$$

From equations 5.19 and 5.20

$$\begin{bmatrix} X \cdot y \\ y \end{bmatrix} F_X = \begin{bmatrix} D_{X^{(2)} F_Y} & D_{X F_Y} \\ D_{X F_Y} & (F_Y' 1) \cdot I \end{bmatrix} - \begin{bmatrix} X \\ 1 \ 1' \end{bmatrix} \frac{D_Y}{n} \begin{bmatrix} X', 1 \ 1' \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} \quad (5.21)$$

where $X^{(2)}$ indicates the elemental exponentiation of X , and the D 's in the first right hand matrix have the same function as in equation 5.19.

To simplify further analysis, we define

$$\gamma \equiv \begin{bmatrix} X \cdot y \\ y \end{bmatrix} F_X \quad (5.22)$$

$$\lambda \equiv \begin{bmatrix} D_{X^{(2)} F_Y} & D_{X F_Y} \\ D_{X F_Y} & (F_Y' 1) \cdot I \end{bmatrix} \quad (5.23)$$

$$d = D_Y / n \quad (5.24)$$

$$v \equiv \begin{bmatrix} X \\ 1 \ 1' \end{bmatrix} \quad (5.25)$$

$$\gamma \equiv v \ d^{\frac{1}{2}} \quad (5.26)$$

$$u \equiv \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (5.27)$$

$$u \equiv \begin{bmatrix} A \\ B \end{bmatrix} \quad (5.28)$$

Substituting equations 5.22 through 5.28 in equation 5.21, we get

$$(\lambda - V V') \mu = \gamma \quad (5.29)$$

Now from equations 5.10 and 5.11, from 5.16, 5.17, 5.18, and 5.20, and from 5.23 through 5.27, it is clear that λ , V , and γ all involve μ . We shall assume that an initial approximation to μ is available from which first approximations to λ , V , and γ may be calculated, and that from these a second approximation to μ may be calculated. Without at this time setting forth computational procedures for these calculations, we shall assume that the cycles may be iterated until the solution for μ stabilizes and θ in equation 5.6 becomes a maximum. Once this stabilization is achieved, we calculate the x matrix given by equation 5.1 and interchange the roles of X and Y . We let

$$y_{.j} \approx x_{.j} / \sigma_{x_j} \quad (5.30)$$

and

$$X \approx Y \quad (5.31)$$

where \approx indicates replacement. We then proceed with the same analysis as before. But now we solve for the A and B row transformation vectors for the original Y matrix which will yield a maximum θ for the calculated x matrix. This second θ should be as great as, or greater than, the first θ that we solved for.

We may thus alternate successively between A and B solutions for X and Y until θ stabilizes.

VI. The Non-Basic Solution

We must now turn our attention to the computational details implied by equation 5.29. It may appear that the solution for any given approximation to μ corresponding to either the X or the Y matrix could be obtained merely by premultiplying γ by the inverse of the parentheses on the left, viz., by $(\lambda - V V')^{-1}$. We shall assume for the present that n , the number of entities,

is much greater than m , the number of variables. The order of λ is $2n$ and the order of V is $2n \times m$. From equation 5.23 λ is seen to be a second order symmetric supermatrix whose submatrix elements are n th order diagonal matrices. The inverse of such a matrix is readily calculated, as we shall show subsequently. Ordinarily we could write

$$(\lambda - V V')^{-1} = \lambda^{-1} + \lambda^{-1} V (I - V' \lambda^{-1} V)^{-1} V \lambda^{-1} \quad (6.1)$$

The parentheses in the right side of equation 6.1 is of order m only which by assumption is much less than $2n$. If λ^{-1} can be readily calculated, then the right side of equation 6.1 suggests a simple way of calculating the inverse of the $2n$ th order parentheses on the right, since only the inverse of an m th order matrix is required. But this suggested solution implies that $(I - V' \lambda^{-1} V)$ has a regular inverse.* Closer examination of the matrix on the left of equation 6.1 will show that it is non-basic and therefore cannot have a regular inverse. But it can be shown that λ is not in general singular and hence can have a regular inverse. If we show that the $(\lambda - V V')$ is non-basic and therefore cannot have a regular inverse, then we must also conclude that the right of equation 6.1 is not a regular inverse solution. But if λ has a regular inverse, the right side of equation 6.1 can fail to be a regular inverse only if $(I - V' \lambda^{-1} V)$ has no regular inverse and equation 6.1 cannot be calculated.

To show that $(\lambda - V V')$ cannot have a regular inverse, it is sufficient to find a vector that is orthogonal to it. We can show that

$$(\lambda - V V') u = 0 \quad (6.2)$$

where u is defined by equation 5.27. The proof follows the same general lines as for equations 2.45 through 2.48.

We shall see now how we may obtain a solution for μ in equation 5.29. From equations 5.29 and 6.2 we have

$$u' \gamma = 0 \quad (6.3)$$

From equation 5.29

$$(I - \lambda^{-1} V V') \mu = \lambda^{-1} \gamma \quad (6.4)$$

From equation 6.3

$$(I - \lambda^{-1} V V') u = 0 \quad (6.5)$$

From equation 6.4

$$(I - V' \lambda^{-1} V) V' \mu = V' \lambda^{-1} \gamma \quad (6.6)$$

From equation 6.5

$$(I - V' \lambda^{-1} V) V' u = 0 \quad (6.7)$$

Let

$$v = V' u (u' V V' u)^{-\frac{1}{2}} \quad (6.8)$$

From equations 6.7 and 6.8

$$(I - V' \lambda^{-1} V) v = 0 \quad (6.9)$$

Write the basic structure form *

$$Q \delta Q' = I - V' \lambda^{-1} V \quad (6.10)$$

From equations 6.6 and 6.10

$$Q \delta Q' V' \mu = V' \lambda^{-1} \gamma \quad (6.11)$$

From equations 6.9 and 6.10

$$Q \delta Q' v = 0 \quad (6.12)$$

From equation 6.11

$$Q' V' \mu = \delta^{-1} Q' V' \lambda^{-1} \gamma \quad (6.13)$$

From equation 6.13

$$V' \mu = Q \delta^{-1} Q' V' \lambda^{-1} \gamma \quad (6.14)$$

From equations 6.10 and 6.12

$$Q \delta^{-1} Q' = (I - V' \lambda^{-1} V + v v')^{-1} - v v' \quad (6.15)$$

From equations 6.14 and 6.15

$$V' \mu = ((I - V' \lambda^{-1} V + v v')^{-1} - v v') V' \lambda^{-1} \gamma \quad (6.16)$$

From equation 6.8

$$v' V' \lambda^{-1} \gamma = (u' V V' u)^{-\frac{1}{2}} u' V V' \lambda^{-1} \gamma \quad (6.17)$$

From equation 6.5

$$u = \lambda^{-1} V V' u \quad (6.18)$$

From equation 6.18

$$u' = u' V V' \lambda^{-1} \quad (6.19)$$

From equations 6.17 and 6.19

$$v' V' \lambda^{-1} \gamma = (u' V V' u)^{-\frac{1}{2}} u' \gamma \quad (6.20)$$

From equations 6.3 and 6.20

$$v' V' \lambda^{-1} \gamma = 0 \quad (6.21)$$

From equations 6.16 and 6.21

$$V' \mu = (I - V' \lambda^{-1} V + v v')^{-1} v' \lambda^{-1} \gamma \quad (6.22)$$

From equations 6.4 and 6.22

$$u - \lambda^{-1} V (I - V' \lambda^{-1} V + v v')^{-1} v' \lambda^{-1} \gamma - \lambda^{-1} \gamma = 0 \quad (6.23)$$

From equation 6.23

$$\mu = (\lambda^{-1} + \lambda^{-1} V (I - V' \lambda^{-1} V + v v')^{-1} v' \lambda^{-1}) \gamma \quad (6.24)$$

From equations 5.26 and 6.8

$$v = d^{\frac{1}{2}} v' u (u' v d v' u)^{-\frac{1}{2}} \quad (6.25)$$

From equations 5.25 and 5.27

$$u' v = n l' \quad (6.26)$$

From equations 6.25 and 6.26

$$v = d^{\frac{1}{2}} n l (n l' d l n)^{-\frac{1}{2}} \quad (6.27)$$

From equation 6.27

$$v = d^{\frac{1}{2}} l (l' d l)^{\frac{1}{2}} \quad (6.28)$$

From equations 5.26, 6.24, and 6.28

$$u = (\lambda^{-1} + \lambda^{-1} v d^{\frac{1}{2}} (I - d^{\frac{1}{2}} v' \lambda^{-1} v d^{\frac{1}{2}} + \frac{d^{\frac{1}{2}} l l' d^{\frac{1}{2}}}{l' d l})^{-1} d^{\frac{1}{2}} v' \lambda^{-1}) \gamma \quad (6.29)$$

From equation 6.29

$$\mu = (\lambda^{-1} + \lambda^{-1} v (d^{-1} - v' \lambda^{-1} v + \frac{l l'}{l' d l})^{-1} v' \lambda^{-1}) \gamma \quad (6.30)$$

VII. Computational Details

Equation 6.30 gives the solution for any given approximation to μ in terms of the immediately preceding approximations to λ , v , and d . We shall therefore examine in detail the computations implied in equation 6.30. First, we shall consider the solution for λ^{-1} . We let

$$D_2 = D_X(2)_{F_Y} \quad (7.1)$$

$$D_1 = D_{XF_Y} \quad (7.2)$$

$$D_0 = (1' D_Y 1) I \quad (7.3)$$

From equations 5.23 and 7.1 through 7.3

$$\lambda \equiv \begin{bmatrix} D_2 & D_1 \\ D_1 & D_0 \end{bmatrix} \quad (7.4)$$

We let

$$\lambda^{-1} \equiv \begin{bmatrix} d_0 & -d_1 \\ -d_1 & d_2 \end{bmatrix} \quad (7.5)$$

and

$$D_\lambda = D_0 D_2 - D_1^2 \quad (7.6)$$

Since the D 's in equation 7.4 are all diagonal matrices, we can readily show that

$$d_0 = D_0 D_\lambda^{-1} \quad (7.7)$$

$$d_1 = D_1 D_\lambda^{-1} \quad (7.8)$$

$$d_2 = D_2 D_\lambda^{-1} \quad (7.9)$$

To develop computational algorithms we shall now expand equation 6.30 in terms of equations 5.22, 5.24, 5.25, 5.28, and 7.5. This gives

$$\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} d_0 & -d_1 \\ -d_1 & d_2 \end{bmatrix} + \begin{bmatrix} d_0 & -d_1 \\ -d_1 & d_2 \end{bmatrix} \begin{bmatrix} X \\ 1 \ 1' \end{bmatrix} \left[n D_Y^{-1} - (X' \ , \ 1 \ 1') \begin{bmatrix} d_0 & -d_1 \\ -d_1 & d_2 \end{bmatrix} \begin{bmatrix} X \\ 1 \ 1' \end{bmatrix} + \right. \\ \left. + \frac{n}{1' D_Y 1} 1 \ 1' \right]^{-1} (X' \ , \ 1 \ 1') \begin{bmatrix} d_0 & -d_1 \\ -d_1 & d_2 \end{bmatrix} \begin{bmatrix} X \cdot y \\ y \end{bmatrix}_{F_Y} \quad (7.10)$$

We now let

$$\begin{bmatrix} X_{YA} \\ X_{YB} \end{bmatrix} = \begin{bmatrix} X'Y \\ Y \end{bmatrix} F_X \quad (7.11)$$

$$\begin{bmatrix} X_A \\ X_B \end{bmatrix} = \begin{bmatrix} d_0 & -d_1 \\ -d_1 & d_2 \end{bmatrix} \begin{bmatrix} X \\ 1 \end{bmatrix} \quad (7.12)$$

$$S = \left[n (D_Y^{-1} + \frac{1 \cdot 1'}{1' D_Y 1}) - (X' , 1 \cdot 1') \begin{bmatrix} X_A \\ X_B \end{bmatrix} \right] \quad (7.13)$$

$$V_Y = (X_A' , X_B') \begin{bmatrix} X_{YA} \\ X_{YB} \end{bmatrix} \quad (7.14)$$

$$S_Y = S^{-1} V_Y \quad (7.15)$$

Then from equations 7.10 through 7.15 we have

$$\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} d_0 & -d_1 \\ -d_1 & d_2 \end{bmatrix} \begin{bmatrix} X_{YA} \\ X_{YB} \end{bmatrix} + \begin{bmatrix} X_A \\ X_B \end{bmatrix} S_Y \quad (7.16)$$

But from equations 5.10, 5.11, 5.16, 5.17, 5.18, and 5.20 it is clear that F_X in equation 7.11 and D_Y in equation 7.13 are themselves functions of A and B .

We therefore start with some first approximation to A and B and calculate all f_{Yj} and f_{Xj} in equations 5.10 and 5.11, respectively. From these we calculate D_X and D_Y by equations 5.16 and 5.17 and construct the vector F_X in equation 5.18. By means of equations 7.1, 7.2, 7.3, and 7.6 through 7.9 we calculate the matrix λ^{-1} in equation 7.5. Then by means of equations 7.11 through 7.16 we calculate a second approximation to A and B . These cycles of iterations continue until A and B stabilize. We also calculate

$$D_r = D_Y D_X \quad (7.17)$$

This by definition gives the correlations between corresponding columns of Y and the row-transformed X . From equations 5.6 and 7.17 we calculate

$$\theta = 1' D_r^{-1} 1 \quad (7.18)$$

This cycle of operations is repeated each time with the last calculated A and B until two successive θ 's are equal to some specified tolerance.

VIII. Supplementary Constraints

It should be noted that θ is independent of any scaling constant we may apply to A_X and B_X or A_Y and B_Y , and also to any additive constant we may apply to B_X or B_Y . We may therefore let

$$a_X = C_X A_X \quad (8.1)$$

$$a_Y = C_Y A_Y \quad (8.2)$$

$$b_X = C_X (B_X + G_X 1) \quad (8.3)$$

$$b_Y = C_Y (B_Y + G_Y 1) \quad (8.4)$$

where C_X and C_Y and also G_X and G_Y may be chosen at will without affecting the solution for θ . We may therefore consider convenient rationales for these scalar constants. We shall distinguish two cases in which the constants may be determined successively.

First, we consider the computational sequences in which we calculate successive approximations to the A_X and A_Y and the B_X and B_Y . Unless certain restrictions are put on these successive solutions computationally, unmanageable drifts may occur. To avoid a complicated investigation of the nature and magnitude of these drifts, we shall simply put constraints of the types 8.1 through 8.4 at each iteration cycle. It is computationally convenient to determine G_X and G_Y in equations 8.3 and 8.4 so that

$$1' b_X = 0 \quad (8.5)$$

$$1' b_Y = 0 \quad (8.6)$$

From equations 8.3 and 8.5 therefore we have

$$G_X = \frac{-1' B_X}{n} \quad (8.7)$$

$$G_Y = \frac{-1' B_Y}{n} \quad (8.8)$$

We may also specify that

$$a_X' a_X + b_X' b_X = k \quad (8.9)$$

$$a_Y' a_Y + b_Y' b_Y = k \quad (8.10)$$

where k remains constant for successive iterations within each cycle for either X or Y and also for successive alternations of cycle sets from X to Y and vice versa. From equations 8.1, 8.3, and 8.9

$$C_X^2 = k / (A_X' A_X + B_X' B_X - n G_X^2) \quad (8.11)$$

and from equations 8.2, 8.4, and 8.10

$$C_Y^2 = k / (A_Y' A_Y + B_Y' B_Y - n G_Y^2) \quad (8.12)$$

A convenient choice for k could be

$$k = 2n \quad (8.13)$$

The second case for determining scaling and additive constants occurs after θ has stabilized. This case does not involve considerations of computational convenience but rather more subjective considerations. It will be recalled that the scaling and additive constants introduced during the iterations are arbitrary. The final row-transformed X and Y matrices based on these arbitrary scalings and origins may bear little or no resemblance to the original X and Y matrices. Without, however, affecting the θ corresponding to the terminal X and Y matrices, one may multiply each matrix by any scalar constant and add any constant to each element of the product. If we let X_T and Y_T be the terminal matrices before adjustment and X_α and Y_α the same matrices after adjustment, we may write

$$X_\alpha = C_X X_T + G_X 1 1' \quad (8.14)$$

$$Y_\alpha = C_Y Y_T + G_Y 1 1' \quad (8.15)$$

Suppose now we let

$$M_{X_\alpha} = 1' X_\alpha 1 / n \quad (8.16)$$

$$M_{Y_{\alpha}} = 1' Y_{\alpha} 1 / n m \quad (8.17)$$

$$\sigma_{X_{\alpha}}^2 = 1' X_{\alpha}^{(2)} 1 / n m - M_{X_{\alpha}}^2 \quad (8.18)$$

$$\sigma_{Y_{\alpha}}^2 = 1' Y_{\alpha}^{(2)} 1 / n m - M_{Y_{\alpha}}^2 \quad (8.19)$$

where the parenthetical exponents on the right of equations 8.18 and 8.19 indicate elemental exponentiation. From the definitions in equations 8.16 through 8.19 we see that $M_{X_{\alpha}}$ and $M_{Y_{\alpha}}$ are the means of all the elements in X_{α} and Y_{α} , respectively, while $\sigma_{X_{\alpha}}$ and $\sigma_{Y_{\alpha}}$ are the corresponding standard deviations. Several reasonable options for choosing these M 's and σ 's suggest themselves.

The simplest options occur when only the X matrix is to be row-transformed. Our data may be experimentally so generated that we may require the M and σ for the adjusted terminal X to be the same as for the original X . Or we may assume that the terminal adjusted X matrix should have the same M and σ as the original Y matrix.

When both the X and Y matrices are row-transformed, we again have two obvious options for adjustment of the terminal X and Y matrices. One of these is to have the adjusted terminal matrices resemble as nearly as possible the corresponding original matrices. In this case, the M and σ for the final X matrix would be the same as for the initial X matrix. This also would be the case with respect to the initial and final Y matrices.

However, we may assume that the original X and Y matrices are both presumed to have been generated in terms of the same natural units and origin. But due to differences in the methods of generation, such as differences in instruction or what not, the original X and Y matrices do not have the same M and σ . If we have no reason to prefer one of the two over the other as a standard, we may take the M and σ of all the elements in both initial matrices. Then both the terminal matrices will be adjusted to yield the M and σ of the com-

bined initial matrices.

Perhaps other criteria for adjusting the terminal matrices might be proposed. In any case, whatever rationales are used for specifying the M 's and σ 's in equations 8.16 through 8.19, we shall now consider the calculation of the C 's and G 's in equations 8.14 and 8.15 to yield the desired M 's and σ 's for the terminal adjusted matrices.

We let

$$M_{X_T} = 1' X_T 1 / n m \quad (8.20)$$

$$M_{Y_T} = 1' Y_T 1 / n m \quad (8.21)$$

$$\sigma_{X_T}^2 = 1' X_T^{(2)} 1 / n m - M_{X_T}^2 \quad (8.22)$$

$$\sigma_{Y_T}^2 = 1' Y_T^{(2)} 1 / n m - M_{Y_T}^2 \quad (8.23)$$

It is then well known that

$$C_X = \frac{\sigma_{X_\alpha}}{\sigma_{X_T}} \quad (8.24)$$

$$C_Y = \frac{\sigma_{Y_\alpha}}{\sigma_{Y_T}} \quad (8.25)$$

$$G_X = M_{X_\alpha} - M_{X_T} C_X \quad (8.26)$$

$$G_Y = M_{Y_\alpha} - M_{Y_T} C_Y \quad (8.27)$$

For any of the four options discussed above we can substitute the corresponding σ_{X_α} , σ_{Y_α} , M_{X_α} , and M_{Y_α} in equations 8.24 through 8.27.

It must be recalled, however, that the A and B vectors calculated so far yielded only X_T and Y_T . Therefore, in order to find the A and B vectors which when applied directly to the initial matrices will yield the final adjusted matrices, we must make corresponding adjustments to the stabilized a and b vectors. Therefore we have

$$A_X = C_X a_X \quad (8.28)$$

$$A_Y = C_Y a_Y \quad (8.29)$$

$$B_X = b_X + G_X 1 \quad (8.30)$$

$$B_Y = b_Y + G_Y 1 \quad (8.31)$$

It must also be noted that the computational algorithms outlined above imply that, in each set of iterations for a new approximation to either a_X , b_X or a_Y , b_Y , these vectors are calculated from a newly row-transformed X or Y as the case may be, and therefore the row transformations to be finally applied to the initial matrices to yield the terminal matrices will be functions of all the intervening a_X , b_X and a_Y , b_Y . This function which must be included in any set of computational algorithms is given by the general recursion formulas for either X or Y by

$$A_{i+1} \approx a_{i+1} \cdot A_i \quad (8.32)$$

$$B_{i+1} \approx a_{i+1} \cdot B_i + b_{i+1} \quad (8.33)$$

where the "dots" on the right of equations 8.32 and 8.33 indicate elemental vector multiplication.

When θ stabilizes for A_{i+1} and B_{i+1} we write, in order to keep the notation consistent with equations 8.28 through 8.31, the general equations for either X or Y as

$$a = A_{i+1} \quad (8.34)$$

$$b = B_{i+1} \quad (8.35)$$

IX. Special Issues

Several points that have not been adequately considered in the foregoing developments should be mentioned. The first concerns the relationship between the number of parameters to be solved for, the number of entities n , and the number of attributes. It can readily be shown that, if the Y matrix is

assumed fixed, then we have $2n$ parameters to be solved for. In general, if this number is equal to the total number of elements in the X matrix, it should be possible to solve for the parameters so that $\theta = 1$ or so that the transformed matrix X would be identically equal to Y . In this case, of course, the solution would not be over-determined and would not be acceptable. Therefore we must have

$$2n < nm$$

or

$$2 < m \quad (9.1)$$

Hence the number of variables must be at least 3 and should presumably be considerably greater.

If we assume transformations in both X and Y , we have $4n$ parameters which should yield a perfect θ with an $n \times m$ matrix. Therefore, to have a scientifically meaningful solution we must have

$$4 < m \quad (9.2)$$

We must for this case have at least 5 and preferably many more variables.

It is interesting to note that both inequalities 9.1 and 9.2 are independent of the number of entities and depend only on the number of variables. So far we have made no attempt to investigate any form of significance test for this model. Presumably, such a test would be a function only of m , θ , and either 2 or 4, as the case may be.

In any event, if we row-transform only the X matrix, it should in general be possible to solve for the transformation vectors such that

$$\theta \geq 2 / m \quad (9.3)$$

If the equality holds, then obviously the θ would not be significant. Similarly, if both matrices were row-transformed, we must have

$$\theta \geq 4 / m \quad (9.4)$$

and if the equality in equation 6.4 holds, clearly θ cannot be significant.

A second point to which we have not addressed our discussion concerns the ratio of n to m . We have assumed in our computational procedures that n is much greater than m . Certainly, experimental situations may well generate data in which the number of variables is much greater than the number of cases. To mention but one, we may have responses of n persons to m test items, each of which may be regarded as a separate variable. In this case the number of items, m , might be much greater than the number of cases, n . Assume now that we also request each person to express on a scale of 0 to 10 his degree of confidence in his answer to each item. Presumably, for such an experimental model one would solve for row transformations of only the confidence matrix. Then the parameters of the solution might be considered as the confidence parameters of the participants.

In any case, the computational procedures would almost certainly be different from those sketched above. These require the inverse solutions of $m \times m$ matrices. If m is much less than n or relatively small, these solutions are feasible. If m is very large, other approaches might be investigated. Before much effort is expended in developing such approaches it is probably well to explore first the extent to which the case for m much greater than n would be meaningful in real world situations.

FOOTNOTE

* Horst, Paul. Matrix algebra for social scientists. New York: Holt, Rinehart and Winston, 1963.

APPENDIX B

OPTIMAL ROW TRANSFORMATIONS OF PAIRED MATRICES: THE SINGLE VECTOR CASE

APPENDIX B

OPTIMAL ROW TRANSFORMATIONS OF PAIRED MATRICES:

THE SINGLE VECTOR CASE

I. The General Case

We have previously considered a model where we have given two $n \times m$ matrices of measures X and Y , purporting to measure the same functions except that, for some or all of the rows of each matrix, linear transformations may have taken place. We sought to determine the parameters of the linear transformations so that the sum of the correlations of corresponding columns of the two transformed matrices would be a maximum. We assumed a set of linear row transformation parameters for X and another set for Y . We had let

X be an $n \times m$ matrix of measures

Y be an $n \times m$ matrix of measures

D_X be an n th order diagonal matrix of scaling constants for rows of X

D_Y be an n th order diagonal matrix of scaling constants for rows of Y

B_X be an n th order vector of additive constants for rows of X

B_Y be an n th order vector of additive constants for rows of Y

$$x = D_X X + B_X \mathbf{1}' \quad (1.1)$$

$$y = D_Y Y + B_Y \mathbf{1}' \quad (1.2)$$

Then x and y are the row-transformed matrices of X and Y , respectively. These transformations are determined so that the sum of correlations of corresponding columns of x and y is

a maximum.

We let

$x_{.j}$ be the j th column vector of x

$y_{.j}$ be the j th column vector of y

$X_{.j}$ and $Y_{.j}$ be similarly defined

$$A_X = D_X 1 \quad (1.3)$$

$$A_Y = D_Y 1 \quad (1.4)$$

$$c_j = x_{.j}' (I - \frac{1 1'}{n}) y_{.j} / n \quad (1.5)$$

$$v_{X_j} = x_{.j}' (I - \frac{1 1'}{n}) x_{.j} / n \quad (1.6)$$

$$v_{Y_j} = y_{.j}' (I - \frac{1 1'}{n}) y_{.j} / n \quad (1.7)$$

$$r_j = \frac{c_j}{v_{X_j}^{\frac{1}{2}} v_{Y_j}^{\frac{1}{2}}} \quad (1.8)$$

$$R_A = \sum r_j / m \quad (1.9)$$

Then A_X , B_X and A_Y , B_Y are to be determined so as to maximize R_A in equation 1.9. Suppose now we define

$$j^{d_{XY}} = \frac{1}{v_{X_j}^{\frac{1}{2}} v_{Y_j}^{\frac{1}{2}}} \quad (1.10)$$

$$j^r = c_j j^{d_{XY}} \quad (1.11)$$

$$j^{d_X} = r_j / v_{X_j} \quad (1.12)$$

$$j^{d_Y} = r_j / v_{Y_j} \quad (1.13)$$

$$d_{XY} = \begin{bmatrix} j^{d_{XY}} & 0 \\ \vdots & \vdots \\ 0 & m^{d_{XY}} \end{bmatrix} \quad (1.14)$$

$$d_X = \begin{bmatrix} d_{X_1} & 0 \\ \vdots & \vdots \\ 0 & d_{X_m} \end{bmatrix} \quad (1.15)$$

$$d_Y = \begin{bmatrix} d_{Y1} & & 0 \\ & \ddots & \\ 0 & & d_{Ym} \end{bmatrix} \quad (1.16)$$

$$F_{XY} = d_{XY} \cdot 1 \quad (1.17)$$

$$F_X = d_X \cdot 1 \quad (1.18)$$

$$F_Y = d_Y \cdot 1 \quad (1.19)$$

$$1' F_X = g_X \quad (1.20)$$

$$1' F_Y = g_Y \quad (1.21)$$

$$1' F_{XY} = h_{XY} \quad (1.22)$$

$$X_1 = X F_X \quad (1.23)$$

$$X_2 = X^{(2)} F_X \quad (1.24)$$

$$Y_1 = Y F_Y \quad (1.25)$$

$$Y_2 = Y^{(2)} F_Y \quad (1.26)$$

where the parenthetical exponents in equations 1.24 and 1.26 indicate elemental exponentiation. Also let

$$C = X \cdot Y F_{XY} \quad (1.27)$$

$$X_Y = X F_{XY} \quad (1.28)$$

$$Y_X = Y F_{XY} \quad (1.29)$$

where $X \cdot Y$ in equation 1.27 indicates elemental multiplication. We further define diagonal matrices as follows:

$$D_{X_1} \cdot 1 = X_1 \quad (1.30)$$

$$D_{X_2} \cdot 1 = X_2 \quad (1.31)$$

$$D_{Y_1} \cdot 1 = Y_1 \quad (1.32)$$

$$D_{Y_2} \cdot 1 = Y_2 \quad (1.33)$$

$$D_C \mathbf{1} = C \quad (1.34)$$

$$D_{X_Y} \mathbf{1} = X_Y \quad (1.35)$$

$$D_{Y_X} \mathbf{1} = Y_X \quad (1.36)$$

Next we construct matrices as follows:

$$\sigma_{XX} = \begin{bmatrix} G_X (I - \frac{\mathbf{1} \mathbf{1}'}{n}) & (I - \frac{\mathbf{1} \mathbf{1}'}{n}) D_{X_1} \\ D_{X_1} (I - \frac{\mathbf{1} \mathbf{1}'}{n}) & D_{X_2} - X \frac{d_X}{n} X' \end{bmatrix} \quad (1.37)$$

$$\sigma_{XY} = \begin{bmatrix} G_{XY} (I - \frac{\mathbf{1} \mathbf{1}'}{n}) & (I - \frac{\mathbf{1} \mathbf{1}'}{n}) D_{Y_X} \\ D_{X_Y} (I - \frac{\mathbf{1} \mathbf{1}'}{n}) & D_C - Y \frac{d_{XY}}{n} X' \end{bmatrix} \quad (1.38)$$

$$\sigma_{YY} = \begin{bmatrix} G_Y (I - \frac{\mathbf{1} \mathbf{1}'}{n}) & (I - \frac{\mathbf{1} \mathbf{1}'}{n}) D_{Y_1} \\ D_{Y_1} (I - \frac{\mathbf{1} \mathbf{1}'}{n}) & D_{Y_2} - Y \frac{d_Y}{n} Y' \end{bmatrix} \quad (1.39)$$

and the vectors

$$W_X' = (B_X', A_X') \quad (1.40)$$

$$W_Y' = (B_Y', A_Y') \quad (1.41)$$

Now in the model referred to above we have shown in somewhat different notation that the solutions for W_X and W_Y which optimize R_A in equation 1.9 can be indicated by

$$\begin{bmatrix} k \sigma_{XX} & - \sigma_{XY} \\ - \sigma_{YX} & k \sigma_{YY} \end{bmatrix} \begin{bmatrix} W_X \\ W_Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (1.42)$$

We noted that equation 1.42 was the well known form for the canonical correlation between two sets of variables. We pointed out that the σ 's were themselves functions of the W 's and that iterative computational procedures were indicated. It was also shown that the left hand bracket in equation 1.42 has a

vanishing root and hence conventional computational procedures required certain modifications. Furthermore, it was found necessary to impose supplementary constraints on the W 's.

II. The Single Vector Case

We shall now consider a modification of the above model. We shall assume that the matrices X and Y have already been subjected to standardizing linear row transformations, so that

$$X \mathbf{1} = 0 \quad (2.1)$$

$$Y \mathbf{1} = 0 \quad (2.2)$$

$$D_{XX}' = m I \quad (2.3)$$

$$D_{YY}' = m I \quad (2.4)$$

Instead of considering two separate sets of transformations W_X and W_Y , we shall attempt to find a single transformation W which, when applied to the row-standardized X and Y matrices, yields the maximum R_A in equation 1.9. Hence we assume that

$$W_X = W_Y = W \quad (2.5)$$

By carrying out the appropriate partial differentiations and equating to zero, the solution for W can be shown to be

$$\left[\bar{\sigma}_{XY} + \sigma_{YX} - k (\sigma_{XX} + \sigma_{YY}) \right] W = 0 \quad (2.6)$$

Suppose we let

$$g = g_X + g_Y \quad (2.7)$$

$$h = 2 h_{XY} \quad (2.8)$$

$$D_a = D_{X_2} + D_{Y_2} \quad (2.9)$$

$$D_b = D_{X_1} + D_{Y_1} \quad (2.10)$$

$$d = \begin{bmatrix} d_X & 0 \\ 0 & d_Y \end{bmatrix} \quad (2.11)$$

$$V = (X, Y) \quad (2.12)$$

$$D_c = 2 D_c \quad (2.13)$$

$$d_f = \begin{bmatrix} 0 & d_{XY} \\ d_{XY} & 0 \end{bmatrix} \quad (2.14)$$

$$q q' = (I - \frac{1}{n} 1 1') \quad (2.15)$$

$$Q = \begin{bmatrix} q & 0 \\ 0 & I \end{bmatrix} \quad (2.16)$$

$$\sigma_1 = \begin{bmatrix} g I & D_b \\ D_b & D_a - V \frac{d}{n} V' \end{bmatrix} \quad (2.17)$$

$$\sigma_2 = \begin{bmatrix} h I & D_e \\ D_e & D_c - V d_f V' \end{bmatrix} \quad (2.18)$$

Then from equations 1.37 through 1.39 and 2.7 through 2.18

$$\sigma_{XX} + \sigma_{YY} = Q Q' \sigma_1 Q Q' \quad (2.19)$$

$$\sigma_{XY} + \sigma_{YX} = Q Q' \sigma_2 Q Q' \quad (2.20)$$

From equations 2.6, 2.19, and 2.20

$$Q Q' (\sigma_2 - k \sigma_1) Q Q' W = 0 \quad (2.21)$$

Now let

$$S_1 = Q' \sigma_1 Q \quad (2.22)$$

$$S_2 = Q' \sigma_2 Q \quad (2.23)$$

$$w = Q' W \quad (2.24)$$

From equations 2.21 through 2.24

$$(S_2 - k S_1) w = 0 \quad (2.25)$$

From equation 2.25

$$(S_1^{-1} S_2 - k I) w = 0 \quad (2.26)$$

From equations 2.17 and 2.22 it can be proved that

$$S_1^{-1} = \begin{bmatrix} q' & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \frac{1}{g} I & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{g} D_b \\ -I \end{bmatrix} \left[D_a - V \frac{d}{n} V' - \frac{1}{g} D_b q q' D_b \right]^{-1} \begin{bmatrix} D_b \frac{1}{g} & -I \end{bmatrix} \begin{bmatrix} q & 0 \\ 0 & I \end{bmatrix} \quad (2.27)$$

But from equation 2.15

$$D_b q q' D_b = D_b \left(I - \frac{1}{n} l l' \right) D_b \quad (2.28)$$

and from equations 1.23, 1.25, and 2.10

$$D_b l = V d l \quad (2.29)$$

From equations 2.28 and 2.29

$$V \frac{d}{n} V' + \frac{1}{g} D_b q q' D_b = D_b^2 / g - V d \frac{1}{g n} l l' d V \quad (2.30)$$

Let

$$D_P = D_a - D_b^2 / g \quad (2.31)$$

$$U = V d^{\frac{1}{2}} \left(I - d^{\frac{1}{2}} \frac{1}{g} l l' d^{\frac{1}{2}} \right) \quad (2.32)$$

From equations 2.30, 2.31, and 2.32

$$D_a - V \frac{d}{n} V' - \frac{1}{g} D_b q q' D_b = D_P - \frac{U U'}{n} \quad (2.33)$$

But

$$(D_P - U U')^{-1} = D_P^{-\frac{1}{2}} \left(I + D_P^{-\frac{1}{2}} U (n I - U' D_P^{-1} U)^{-1} U' D_P^{-\frac{1}{2}} \right) D_P^{-\frac{1}{2}} \quad (2.34)$$

Let

$$u = D_P^{-\frac{1}{2}} U \quad (2.35)$$

$$S_u = n I - u' u \quad (2.36)$$

Let

$$S_3 = \begin{bmatrix} \frac{1}{g} I & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{g} D_b \\ -I \end{bmatrix} D_P^{-1} \left[I + u S_u^{-1} u' \right] D_P^{-1} \begin{bmatrix} D_b \frac{1}{g} & -I \end{bmatrix} \quad (2.37)$$

From equations 2.16 and 2.27 through 2.37 we have, finally,

$$S_1^{-1} = Q' S_3 Q \quad (2.38)$$

From equations 2.23 through 2.26 and 2.38 we have

$$(Q' S_3 Q Q' \sigma_2 Q - k I) Q' W = 0 \quad (2.39)$$

Let

$$t = Q' Q' W \quad (2.40)$$

From equations 2.39 and 2.40

$$Q Q' S_3 Q Q' \sigma_2 t = k t \quad (2.41)$$

It should be noted that the effect of the $Q Q'$ matrix is to deviate the B part of the W vector.

III. Computational Procedure

The solution for the t vector in equation 2.41 is iterative. We begin by standardizing the X and Y matrices by rows. Then we take some approximations to the B and A vectors such that

$$B' 1 = 0 \quad (3.1)$$

We may also specify that

$$A' A = n \quad (3.2)$$

Obvious first approximations to B and A could be

$$B = 0 \quad (3.3)$$

$$A = 1 \quad (3.4)$$

We next calculate

$$X_{ij} = A_i X_{ij} + B_i \quad (3.5)$$

$$Y_{ij} = A_i Y_{ij} + B_i \quad (3.6)$$

$$X_{1j} = \sum_i X_{ij} \quad (3.7)$$

$$X_{2j} = \sum_i X_{ij}^2 \quad (3.8)$$

$$Y_{1j} = \sum_i Y_{ij} \quad (3.9)$$

$$Y_{2j} = \sum_i Y_{ij}^2 \quad (3.10)$$

$$R_j = \sum_i X_{ij} Y_{ij} \quad (3.11)$$

Then we calculate the replacement values

$$X_{1j} = X_{1j} / n \quad (3.12)$$

$$Y_{1j} = Y_{1j} / n \quad (3.13)$$

$$X_{2j} = X_{2j} / n - X_{1j}^2 \quad (3.14)$$

$$Y_{2j} = Y_{2j} / n - Y_{1j}^2 \quad (3.15)$$

and the values

$$F_{XYj} = 1 / (X_{2j} Y_{2j})^{\frac{1}{2}} \quad (3.16)$$

$$R_j = (R_j / n - X_{1j} Y_{1j}) F_{XYj} \quad (3.17)$$

$$F_{Xj} = R_j / X_{2j} \quad (3.18)$$

$$F_{Yj} = R_j / Y_{2j} \quad (3.19)$$

$$G = 1' F_X + 1' F_Y \quad (3.20)$$

$$H = 2 1' F_{XY} \quad (3.21)$$

$$R_A = 1' R / n \quad (3.22)$$

Next we calculate the D values in equations 2.17 and 2.18 by

$$D_a 1 = X^{(2)} F_X + Y^{(2)} F_Y \quad (3.23)$$

$$D_b 1 = X F_X + Y F_Y \quad (3.24)$$

$$D_c 1 = 2 (X \cdot Y) F_{XY} \quad (3.25)$$

$$D_e 1 = X F_{XY} + Y F_{XY} \quad (3.26)$$

and the replacement equations

$$D_a l = (D_a l - D_b^2 l / G)^{(-\frac{1}{2})} \quad (3.27)$$

$$F_X = F_X^{(\frac{1}{2})} \quad (3.28)$$

$$F_Y = F_Y^{(\frac{1}{2})} \quad (3.29)$$

At this time we also calculate the u matrices in equation 2.35 by

$$u_X = D_a^{-1} (X - D_b \frac{l l'}{G}) D_{F_X} \quad (3.30)$$

$$u_Y = D_a^{-1} (Y - D_b \frac{l l'}{G}) D_{F_Y} \quad (3.31)$$

To calculate $\sigma_2 t$ in equation 2.41 we write from equation 2.18

$$\sigma_2 t = \begin{bmatrix} h I & D_e \\ D_e & D_c - V d_f V' \end{bmatrix} \begin{bmatrix} B \\ A \end{bmatrix} \quad (3.32)$$

We let

$$A_a = (D_e B + D_c - V d_f V') A \quad (3.33)$$

First we have

$$V' A = \begin{bmatrix} X' \\ Y' \end{bmatrix} A \quad (3.34)$$

Hence we calculate

$$X_A = X' A \quad (3.35)$$

$$Y_A = Y' A \quad (3.36)$$

Then

$$X_A = D_{F_{XY}} X_A \quad (3.37)$$

$$Y_A = D_{F_{XY}} Y_A \quad (3.38)$$

Next we calculate

$$B_b = h B + D_e A \quad (3.39)$$

$$A_a = D_e B + D_b A \quad (3.40)$$

$$A_a = A_a - X Y_A - Y X_A \quad (3.41)$$

and finally

$$B_b = (I - \frac{1}{n} \frac{1'}{n}) B_b \quad (3.42)$$

We now have

$$Q Q' \sigma_2 t = \begin{bmatrix} B_b \\ A_a \end{bmatrix} \quad (3.43)$$

Next we shall calculate the S_u matrix in equation 2.36. From the matrices in equations 3.30 and 3.31 we calculate

$$S_u = \begin{bmatrix} n I & 0 \\ 0 & n I \end{bmatrix} - \begin{bmatrix} u_X' \\ u_Y' \end{bmatrix} \begin{bmatrix} u_X & u_Y \end{bmatrix} \quad (3.44)$$

We calculate first the negative of the upper triangular part of $u_X' u_X$, then the negative of the $u_X' u_Y$ submatrix and of the upper triangular part of $u_Y' u_Y$. The lower part of the supermatrix is then filled in and n is added to the diagonals. The inverse of S_u is next calculated by a standard routine.

We are now ready to premultiply equation 3.43 by S_3 as indicated in equation 2.41. We first calculate

$$B_{bb} = D_a (D_b B_b / G - A_a) \quad (3.45)$$

Next we calculate $u' B_{bb}$ as

$$B_X = u_X' B_{bb} \quad (3.46)$$

$$B_Y = u_Y' B_{bb} \quad (3.47)$$

Indicating

$$S_u^{-1} = \begin{bmatrix} S_{XX} & S_{XY} \\ S_{YX} & S_{YY} \end{bmatrix} \quad (3.48)$$

we calculate

$$B_{XX} = S_{XX} B_X + S_{XY} B_Y \quad (3.49)$$

$$B_{YY} = S_{YX} B_X + S_{YY} B_Y \quad (3.50)$$

Next we calculate

$$B_{bb} = B_{bb} + u_X B_{XX} + u_Y B_{YY} \quad (3.51)$$

The next approximation to A is given by

$$A = -D_a B_{bb} \quad (3.52)$$

and to the uncentered B by

$$B = (B_b + D_b A) / g \quad (3.53)$$

We then deviate B by

$$B = (I - \frac{1 \cdot 1'}{n}) B \quad (3.54)$$

Also we rescale A and B so that

$$A = A \left(\frac{n}{A' A} \right)^{\frac{1}{2}} \quad (3.55)$$

$$B = B \left(\frac{n}{A' A} \right)^{\frac{1}{2}} \quad (3.56)$$

We now return to equation 3.5 with the new values of B and A and proceed through equation 3.22, the average column correlation. If this is not greater than the previous R_A by some specified tolerance, the calculations cease. Otherwise they continue until a specified tolerance or iteration limit is reached.

Suitable further transformations may be worked out so that the terminal X and Y matrices satisfy prespecified conditions. Such transformations must of course preserve the initial

row correlations and the final column correlations. Presumably they would therefore be restricted to linear transformations of all elements in the X and Y matrices.

Furthermore, it might be desirable to impose special constraints on the A and B vectors. For example, it might be required that no element of the A vector be less than some prespecified positive value.

APPENDIX C

A SIMPLIFIED MODEL FOR OPTIMAL ROW TRANSFORMATIONS OF PAIRED MATRICES

APPENDIX C

A SIMPLIFIED MODEL FOR OPTIMAL ROW TRANSFORMATIONS OF PAIRED MATRICES

We have previously considered a model where we have given two $n \times m$ matrices of measures X and Y . We assumed that they purport to measure the same function except that, for some or all of the rows of each matrix, unknown linear transformations may have taken place. We sought to determine the parameters of the linear transformations so that the sum of the correlations of corresponding columns of the two transformed matrices would be a maximum. We let A_X , B_X and A_Y , B_Y be respectively the vectors of proportionality constants for the rows of the corresponding X and Y matrices. The solution considered restrictions on the A vectors such that no element of A_X or A_Y was less than some prespecified proportion of the highest element. Further, the sum of the elements in each B vector was restricted to zero. Without restrictions on the A and B vectors it was found that the solution yielded unrealistically high values on a pair of experimental data matrices and also on data matrices generated from random numbers.

Other models than the one above may also be suggested for pairs of data matrices. For example, instead of assuming independent linear row transformations for each matrix, one may assume the same set of linear transformations applied to both matrices. If, however, the quantification strategies are

different for the X and Y matrices, one may first wish to reduce both to the same mean and standard deviation for all elements before considering the common linear row transformations. It appears that the computational algorithms required for determining the transformation vectors which will maximize the sum of corresponding column correlations may be rather complicated. The procedure involves setting up the criterion function involving the unknown A and B vectors. The first derivatives of the function with respect to the vectors are equated to zero. These derivatives are themselves non linear functions of the elements of A and B and the solution involves successive approximations. Problems of convergence and local optima are encountered which further complicate the numerical solution.

We may, however, consider another model which, though somewhat less sophisticated mathematically than the one just suggested, is more feasible computationally, and intuitively, at least, may have some theoretical justification. As with the two models considered above, we begin with two $n \times m$ data matrices X and Y obtained on the same group of n entities. We assume that the matrices purport to measure more or less closely the same m functions. However, the origin and unit of measurement for one matrix may not be the same as for the other, due to differences in methods of data generation and quantification. In any case, we shall consider only row linear transformations so that the correlations for corresponding rows of the two matrices will remain invariant.

Briefly, the suggested procedure is to begin by normalizing rows of both the X and Y matrices. We may indicate these operations by

$$x = D_X X - B_X l' \quad (1)$$

$$y = D_Y Y - B_Y l' \quad (2)$$

where D_X and D_Y are diagonal and B_X and B_Y are vectors and where these are determined such that

$$x l = 0 \quad (3)$$

$$y l = 0 \quad (4)$$

$$D_{xx'} = m I \quad (5)$$

$$D_{yy'} = m I \quad (6)$$

Equations 3 and 4 indicate that the row sums of x and y are zero, and equations 5 and 6 indicate that the corresponding row standard deviations are all unity.

We next consider a row scaling of both x and y by the same diagonal matrix D which we indicate by

$$u = D x \quad (7)$$

$$v = D y \quad (8)$$

Before specifying the matrix D , we first represent a standardization of the u and v matrices by

$$w = (I - \frac{1 l'}{n}) u d_u \quad (9)$$

$$z = (I - \frac{1 l'}{n}) v d_v \quad (10)$$

where d_u and d_v are diagonal matrices of the reciprocals

of the standard deviations of the columns of u and v , respectively. Our interest now is directed to the correlations between corresponding columns of w and z . The diagonal of these correlations can be indicated by

$$D_r = D_w' z / n \quad (11)$$

where the right side of equation 11 is a diagonal of the average cross product for corresponding columns of w and z . These diagonal elements are by definition correlation coefficients.

We could now write the function

$$\phi = \text{tr } w' z / n \quad (12)$$

which is by definition the sum of corresponding column correlations of w and z and is a function of the elements of D in equations 7 and 8. We could then indicate the derivative of ϕ with respect to D and equate to 0 thus:

$$\frac{d \phi}{d D} = 0 \quad (13)$$

Equation 13 would yield a somewhat complicated nonlinear function of the unknown elements of D which could presumably be solved for by appropriate iterative procedures.

However, we might adopt an approach that is somewhat less sophisticated mathematically for the determination of D . We begin by considering the special case where the row correlations between X and Y are all unity. In this case then, the row correlations between x and y in equations 1 and 2 would also be unity and we would have identically, because of equations 3 through 6, that

$$x = y \quad (14)$$

Because of equations 7 through 10 we would also have

$$w = z \quad (15)$$

and therefore equation 11 becomes

$$I = D_{w'z} / n \quad (16)$$

or all the column correlations between w and z are also unity. Therefore we can take D in equations 7 and 8 as

$$D = I \quad (17)$$

Now if all the row correlations between X and Y are not unity, we might with some justification argue that corresponding rows of x and y in equations 7 and 8 should be given a weight which is a monotonically increasing function of the corresponding row correlations. In the special case of equations 14 and 15, D would then be given as equation 16. In general, we can write

$$D = f(D_r) \quad (18)$$

where the first derivative of equation 18 with respect to D_r is positive, thus:

$$\frac{d D}{d D_r} > 0 \quad (19)$$

In particular, we could have

$$D = k D_r \quad (20)$$

where k is some positive constant. But equation 20 might involve difficulties if some of the r 's in D_r are negative. Therefore it may be more appropriate to write

$$D = A + B D_r \quad (21)$$

where A and B are constants to be determined. Suppose, for example, that the highest r in D_r is r_H and the smallest is r_S . We may wish to require that the highest D value be 1 and the lowest be some percentage value P . It can readily be shown that for A and B , respectively, in equation 21 we would have

$$A = (P r_H - r_S) / (r_H - r_S) \quad (22)$$

$$B = (1 - P) / (r_H - r_S) \quad (23)$$

We may, however, instead of equation 21, argue that all cases whose corresponding r values are less algebraically than some percentage value P of the highest correlation r_H are defective cases that did not respond rationally or appropriately to the data generating procedures, and hence should be given minimal weight in determining the column correlations. Thus, for example, all values of $P r_H$ or less may be given a weight of P_S where $0 \leq P_S \leq P r_H$. Other analytic or non-analytic functions in equation 18 may be formulated and tried out with experimental and fictitious data.

Although the model outlined here lacks the mathematical sophistication of the two considered at the beginning of this discussion, it may have at least two advantages. First, its rationale and computational procedures are uncomplicated. Second, degrees of freedom are not sacrificed in the solution for the transformation vectors A and B as they are necessarily when loss functions are minimized.

In order to find the transformations which must be applied to the original X and Y matrices to yield the u and v

matrices, let

X_1 = a vector of X row means

X_2 = a vector of X row sigmas

Y_1 = a vector of Y row means

Y_2 = a vector of Y row sigmas

A_X = a vector of X scaling constants

A_Y = a vector of Y scaling constants

B_X = a vector of X additive constants

B_Y = a vector of Y additive constants

Then

$$A_X = (A + B D_r 1) / X_2 \quad (24)$$

$$A_Y = (A + B D_r 1) / Y_2 \quad (25)$$

$$B_X = A_X \cdot X_1 \quad (26)$$

$$B_Y = A_Y \cdot Y_1 \quad (27)$$

where the dot in the last two equations means elemental multiplications. These vectors when applied to the corresponding X and Y matrices yield the u and v matrices whose corresponding column correlations are to be calculated.

APPENDIX D
SIMILARITY TRANSFORMATIONS FOR PAIRED MATRICES

APPENDIX D

SIMILARITY TRANSFORMATIONS FOR PAIRED MATRICES

Suppose we have given two $n \times m$ matrices of measures X_1 and X_2 for the same n entities on the same m variables. In general, the measures for the two matrices may have been obtained by different experimental methods, different quantification strategies, and at different times. We wish to consider certain transformations of each matrix which will yield new matrices u_1 and u_2 and which in some sense are similar. These transformations shall satisfy the following conditions:

1. The vector of row means of u_1 shall be the same as the vector of row means of u_2 .
2. The vector of row standard deviations of u_1 shall be the same as the vector of row standard deviations of u_2 .
3. The vector of correlations of corresponding rows of u_1 and u_2 shall be the same as the vector of correlations of corresponding rows of X_1 and X_2 .

These transformations will be taken in steps of three sets.

I. The Total Matrix Standardization

First, we shall multiply each element of X_1 by a constant A_1 and subtract a constant B_1 from each element to get a new matrix x_1 . The constants A_1 and B_1 are determined so that the mean of all the elements in the new matrix x_1 is zero and the standard deviation of all the elements is unity.

Similarly, we transform each element of X_2 by a multiplying constant A_2 and an additive constant B_2 to get a new matrix x_2 . These constants are determined so that the mean of all the elements in x_2 is zero and the standard deviation of all the elements is unity.

The new matrices x_1 and x_2 will satisfy the following conditions:

1. The n th order entity-by-entity deviation covariance matrix of x_1 will be the same as for X_1 except for the proportionality constant A_1 .
2. The m th order attribute-by-attribute deviation covariance matrix of x_1 will be the same as for X_1 except for the proportionality constant A_1 .

3. Conditions (1) and (2) will also be satisfied by x_2 with respect to X_2 except for the proportionality constant A_2 .

4. The n th order entity-by-entity deviation covariance matrix of x_1 with x_2 will be the same as for X_1 with X_2 except for the proportionality constant $A = A_1 \times A_2$.

5. The m th order attribute-by-attribute deviation covariance matrix of x_1 with x_2 will be the same as for X_1 with X_2 except for the proportionality constant $A_1 A_2$.

6. The n th order entity-by-entity correlation matrix of x_1 is exactly the same as for X_1 .

7. The m th order attribute-by-attribute correlation matrix of x_1 is exactly the same as for X_1 .

8. Conditions (6) and (7) will also be satisfied by x_2 with respect to X_2 .

9. The n th order entity-by-entity correlation matrix of x_1 with x_2 will be exactly the same as for X_1 with X_2 .

10. The m th order attribute-by-attribute correlation matrix of x_1 with x_2 will be exactly the same as for X_1 with X_2 .

11. The sum of the row means of x_1 is zero.

12. The sum of the column means of x_1 is zero.

13. Conditions (11) and (12) will also be satisfied for x_2 .

14. Condition (1) implies that the vector of row variances of x_1 is the same as the vector of row variances of X_1 except for the proportionality constant A_1 .

15. Condition (2) implies that the vector of column variances of x_1 is the same as the vector of column variances of X_1 except for the proportionality constant A_1 .

16. Condition (3) implies that conditions (14) and (15) will be satisfied, except for the proportionality constant A_2 , for x_2 and X_2 .

17. Condition (4) implies that the vector of covariances of corresponding rows of x_1 and x_2 will be the same as for X_1 and X_2 except for the constant $A_1 A_2$.

18. Condition (5) implies that the vector of covariances of corresponding columns of x_1 with x_2 will be the same as for X_1 and X_2 except for the constant $A_1 A_2$.

These conditions are spelled out in considerable detail to show

that we have lost no essential information in the type I transformations of the X to x matrices. At the same time, the latter are more nearly comparable. As a matter of fact, x_1 and x_2 are the same, except for scale and origin of measurement, as X_1 and X_2 .

We shall therefore consider x_1 and x_2 as the matrices of primary interest rather than X_1 and X_2 , since the former can be more simply and meaningfully discussed. We shall consider further operations on these matrices which will satisfy the three minimal conditions to be satisfied by the final u_1 and u_2 matrices. We therefore consider the second set of transformations which we shall apply to the x_1 and x_2 matrices.

II. The Row Standardization Matrices

We consider a multiplying column vector D_1 and an additive vector b_1 which, when applied to the rows of x_1 , yield a matrix Z_1 . Similarly, we consider D_2 and b_2 vectors which, when applied to the rows of the x_2 matrix, yield a Z_2 matrix. The D and b vectors are determined so that the Z matrices have the following properties:

1. The column vector of row means of each Z matrix is zero.
2. The column vector of row variances of each matrix is a unit vector.
3. The column vector of corresponding row correlations of Z_1 with Z_2 is the same as the vector of correlations of corresponding rows of x_1 with x_2 .
4. The sum of column means for each of Z_1 and Z_2 is zero.

We see therefore that the conditions (1) and (2) satisfied by the Z_1 and Z_2 matrices are special cases of the first two conditions to be satisfied by the u_1 and u_2 matrices. The third condition satisfied by the Z_1 and Z_2 matrices is exactly the third condition to be satisfied by the u_1 and u_2 matrices. But this second set of transformations is more restrictive than necessary to satisfy the first two conditions specified for u_1 and u_2 . They have been introduced to facilitate further analytical development of the model. We therefore consider further row linear transformations of the Z_1 and Z_2 matrices.

III. Row Linear Transformations

We consider a single multiplying column vector D and a single column additive vector B to be applied to the rows of both Z_1 and Z_2 . These resulting matrices are the u_1 and u_2 matrices which we set out to develop. We shall, however, constrain the D and B vectors so that the u matrices satisfy the following conditions:

1. The sum of all the elements in each of u_1 and u_2 is zero.
2. The standard deviation of all the elements in each of u_1 and u_2 is unity.

These constraints can be imposed without loss of generality, just as in the case of the x_1 and x_2 matrices.

It should be noted that all three sets of transformations could have been combined into a single set of linear row transformations for the original X_1 and X_2 matrices to yield directly the u_1 and u_2 matrices. However, the three-stage breakdown of the sets of transformations simplifies further analysis.

The u_1 and u_2 matrices now satisfy the three conditions we have specified that they satisfy. We shall now list certain important relationships involving the D and B vectors which are either self-evident or will be proved in a later section.

1. The vector of row means for both the u_1 and u_2 matrices is the B vector.
2. The vector of row standard deviations for both the u_1 and u_2 matrices is the D vector.
3. The sum of row means for each u matrix is zero.
4. The sum of column means for each u matrix is zero.
5. The sum of the B or additive elements is zero.
6. The average of the row variances for each u matrix is equal to the average of the squared elements in the D scaling vector.
7. The average of the squares of the scaling (D) and additive (B) vector elements is unity.
8. The average of the column variances for both u 's is equal to unity less the variance of their column means.
9. The variance of row means for both u_1 and u_2 is the variance of elements in the additive vector B .
10. The average of corresponding column covariances for u_1 and u_2 is equal to the average of row covariances for corresponding rows

of u_1 and u_2 , plus the variance of row means, less the covariance of their column means.

11. The average row covariances for corresponding rows of u_1 and u_2 are equal to the average product of row correlations by row variances.

12. The larger the variance of row means the larger the average covariance of corresponding columns of u_1 with u_2 .

13. The larger the variance of the additive vector B , the smaller will be the average of the column variances for either u_1 or u_2 .

14. If the number of entities is greater than the number of attributes, the D vector can be chosen so that the corresponding column means of u_1 and u_2 are equal.

15. If the corresponding column means of u_1 and u_2 are equal, the average column variances of u_1 and u_2 are equal.

16. If the variance of the row means, i.e., the variance of the B vector, is chosen to be equal to the variance of the column mean of u_1 and u_2 , then the average column variance will be equal to the average row variance.

17. If the average row and column variances for u_1 and u_2 are all equal, then the average of corresponding column covariances for the two matrices will be equal to the average of corresponding row covariances.

18. The higher the correlation of the squared D elements with the row correlations of u_1 with u_2 , the higher the average column covariances of u_1 with u_2 .

19. If the average column variances for u_1 and u_2 are all equal, then the sums of the elements in the $(m \times m)$ u_1 covariance matrix, the u_2 covariance matrix, and the covariance matrix of u_1 with u_2 are all zero.

We consider now a number of possible restrictions which we may put on D and B .

1. D and B may be proportional to the row standard deviations and means of x_1 .

2. D and B may be proportional to the row standard deviations and means of x_2 .

3. Conditions (1) and (2) may be varied so that a different proportionality constant applies to D than to B . In any case, the average square of the D plus the B elements should be unity.

4. If the number of entities (n) is equal to or greater than the number of attributes (m), D may be determined so that the column means of u_1 are equal to the corresponding means of u_2 .

5. If n is greater than m , we may require that the elements of D be as nearly equal as possible.

6. If n is greater than or equal to m , we may require that the column variances of u_1 be equal to the corresponding column variances of u_2 .

7. If n is greater than m , we may also require that the variance of the elements in the B vector is a minimum.

8. If n is greater than m , B may be determined so that the average column and row variances of both u_1 and u_2 are equal and the individual column variances are all as nearly equal as possible.

9. If n is greater than m , we determine B so that the average column and row variances of both u_1 and u_2 are all equal and the column variances of u_1 are as nearly equal to the corresponding column variances of u_2 as possible.

10. If m is greater than n , then the restriction that corresponding column means of the two matrices shall be as nearly equal as possible, and corresponding column variances shall also be as nearly equal as possible, completely determines the D and B vectors.

The restrictions in (1) and (2) above have been utilized by Nebeker(1976) and those in (4) through (8) by Horst.* The analyses utilizing these restrictions were applied to a pair of 128 x 7 data matrices.

Other variations of the row transformation models may be developed which seek to maximize corresponding column covariances and/or correlations. As earlier research has indicated, these models require appropriate constraints on the D and B vectors but further experimentation with them should probably be considered.

Approaches not restricted to linear row transformations might also be profitable for further study and research.

* See Appendix G.

APPENDIX E
THE WEIGHTED AVERAGE ROW SUM AND VARIANCE MODEL

APPENDIX E

THE WEIGHTED AVERAGE ROW SUM AND VARIANCE MODEL

Given the matrices x_1 and x_2 such that

$$\sum_j x_{ij} = 0 \quad (1)$$

$$\sum_j x_{ij} = 0 \quad (2)$$

$$\sum_j x_{ij}^2 = n \quad (3)$$

$$\sum_j x_{ij}^2 = n \quad (4)$$

Let x be a weighted average of x_1 and x_2 so that

$$P_1 x_1 + P_2 x_2 = x \quad (5)$$

where

$$P_1 + P_2 = 1 \quad (6)$$

Let

M_x be a column of row means of x

D_x^2 be a diagonal of row variances of x

M_{x_1} be a column of row means of x_1

M_{x_2} be a column of row means of x_2

$D_{x_1}^2$ be a diagonal of row variances of x_1

$D_{x_2}^2$ be a diagonal of row variances of x_2

D_r be a diagonal of row correlations between x_1 and x_2

It can be shown then that

$$M_x = P_1 M_{x_1} + P_2 M_{x_2} \quad (7)$$

$$D_x^2 = P_1^2 D_{x_1}^2 + 2 P_1 P_2 D_{x_1} D_{x_2} D_r + P_2^2 D_{x_2}^2 \quad (8)$$

We now consider the row transformations

$$y_1 = D_1 x_1 + B_1 1' \quad (9)$$

$$y_2 = D_2 x_2 + B_2 1' \quad (10)$$

We let M_{y_1} and M_{y_2} be the row means of y_1 and y_2 and $D_{y_1}^2$, $D_{y_2}^2$ be the row variances. We determine D_1 , D_2 and B_1 , B_2 so that

$$M_{y_1} = M_{y_2} = M_x \quad (11)$$

$$D_{y_1} = D_{y_2} = D_x \quad (12)$$

In particular then, if we let $P_1 = 1$, we have $P_2 = 0$ so that

$$M_{y_1} = M_{y_2} = M_{x_1} \quad (13)$$

$$D_{y_1} = D_{y_2} = D_{x_1} \quad (14)$$

If we let $P_2 = 1$, then $P_1 = 0$ and we have

$$M_{y_1} = M_{y_2} = M_{x_2} \quad (15)$$

$$D_{y_1} = D_{y_2} = D_{x_2} \quad (16)$$

To choose a value of P we may consider the covariance matrices:

$$C_1 = y_1' (I - \frac{1 1'}{n}) y_1 / n \quad (17)$$

$$G_1 = y_1 (I - \frac{1 1'}{m}) y_1' / m \quad (18)$$

$$C_2 = y_2' \left(I - \frac{1 \cdot 1'}{n} \right) y_2 / n \quad (19)$$

$$G_2 = y_2 \left(I - \frac{1 \cdot 1'}{m} \right) y_2' / m \quad (20)$$

From equations 17 through 20

$\text{tr } C_1 / m$ is the average of column variances of y_1

$\text{tr } G_1 / n$ is the average of row variances for y_1

$\text{tr } C_2 / m$ is the average of column variances for y_2

$\text{tr } G_2 / n$ is the average of row variances for y_2

If we have some theory as to what conditions should be approximated by the transformed matrix, we may vary the value of P to see what value best satisfies these conditions. For example, it may be desired that the average of the sums of squares of column means for y_1 and y_2 separately should be as nearly equal to each other and to the average of sums of squares of row means as possible.

APPENDIX F

THE PAIRED MATRIX MODEL WITH EQUAL ROW AND COLUMN AVERAGE VARIANCES

APPENDIX F

THE PAIRED MATRIX MODEL WITH EQUAL ROW AND COLUMN AVERAGE VARIANCES

Suppose we have given the two $n \times m$ matrices X_1 and X_2 . We wish to determine transformations of the matrices X_1 and X_2 to u_1 and u_2 such that:

1. The row correlations for u_1 and u_2 are the same as for X_1 and X_2 .
2. The row means for u_1 and u_2 are equal.
3. The row variances for u_1 and u_2 are equal.
4. The average column variances for u_1 and u_2 are equal to the average row variances. These conditions can be satisfied identically if the number of entities is equal to or greater than the number of variables.

If the above four conditions are satisfied, it can be shown that:

5. The averages of squared column means for u_1 and u_2 are equal and equal to the average of squared row means.
6. The average of column covariances for u_1 and u_2 is the same as the average of row covariances.

To put it another way, if conditions 1, 2, 3, and 5 are satisfied, then conditions 4 and 6 will also be satisfied.

The mathematical proofs are developed in the following.

Equations 3 and 4 below mean that the B 's in equations 1 and 2 have been determined so that the row means are all zero.

Equations 5 and 6 mean that the D 's in equations 1 and 2 have been determined so that the row standard deviations are all unity. Therefore x_1 and x_2 are the row-standardized transformations of X_1 and X_2 , respectively.

Equations 7 and 8 mean that x_1 and x_2 have been further transformed so that u_1 and u_2 have the same row means B and standard deviations D .

Equations 9 and 10 are the column sums b_1 and b_2 of u_1 and u_2 , respectively.

Equation 11 indicates that the sums of row means for both u_1 and u_2 are zero.

Equation 12 is simply a vector of row standard deviations.

Let

$$x_1 = D_1 X_1 + B_1 l' \quad (1)$$

$$x_2 = D_2 X_2 + B_2 l' \quad (2)$$

$$x_1 l = 0 \quad (3)$$

$$x_2 l = 0 \quad (4)$$

$$D_{x_1 x_1}' = m I \quad (5)$$

$$D_{x_2 x_2}' = m I \quad (6)$$

$$u_1 = D x_1 + B l' \quad (7)$$

$$u_2 = D x_2 + B l' \quad (8)$$

$$b_1' = l' u_1 \quad (9)$$

$$b_2' = l' u_2 \quad (10)$$

$$B' l = 0 \quad (11)$$

$$V = D l \quad (12)$$

$$b_1 - b_2 = b \quad (13)$$

$$x_1 - x_2 = x \quad (14)$$

From equations 7, 9, 11, and 12

$$b_1 = x_1' V \quad (15)$$

Similarly

$$b_2 = x_2' V \quad (16)$$

From equations 14, 15, and 16

$$b = x' V \quad (17)$$

First let us require that

$$b' b = \phi = \min. \quad (18)$$

with the constraint

$$V' V = k \quad (19)$$

We write

$$\theta = \phi + V' V k \quad (20)$$

$$\frac{\partial \theta}{\partial V'} = 0 \quad (21)$$

From equations 20 and 21

$$((x x') - \phi I) V = 0 \quad (22)$$

Let

$$x = P \delta Q' \quad (23)$$

From equations 22 and 23

$$(P \delta^2 P' - \phi I) V = 0 \quad (24)$$

where, because of equations 3, 4, 14, and 23, P is $n \times (m - 1)$.

If $n > (m - 1)$, we let

$$H \equiv (P, p) \quad (25)$$

where

$$H' H = H H' = I \quad (26)$$

Then we let

$$V = p v \quad (27)$$

where v is a vector of order $n - m + 1$. Because of equations 25 and 26

$$b = 0 \quad (28)$$

To determine v , we may let

$$V - 1 = \epsilon \quad (29)$$

$$\epsilon' \epsilon = \min. \quad (30)$$

From equations 27 and 29 we have

$$\epsilon' \epsilon = v' v - 2 v' p' 1 + n \quad (31)$$

Let

$$\frac{\partial (\epsilon' \epsilon)}{\partial v'} = 0 \quad (32)$$

From equations 31 and 32

$$v = p' 1 \quad (33)$$

From equations 27 and 33

$$V = p p' 1 \quad (34)$$

From equations 25, 26, and 34

$$V = (I - P P')^{-1} 1 \quad (35)$$

To solve for $P P'$ in equation 35 we write

$$x = P Q' \quad (36)$$

Because of equations 3 and 4, Q' is of order $(m-1) \times m$. We let $x_{(j)}$ be x with the j th column deleted and $Q_{(j)}'$ be Q' with the j th row of Q deleted, so that

$$x_{(j)} = P Q_{(j)}' \quad (37)$$

From equation 37, since $Q_{(j)}$ is square, we can readily prove that

$$P P' = x_{(j)} (x_{(j)}' x_{(j)})^{-1} x_{(j)}' \quad (38)$$

From equations 35 and 38

$$V = (I - x_{(j)} (x_{(j)}' x_{(j)})^{-1} x_{(j)}')^{-1} 1 \quad (39)$$

If $n < m$, then ϕ is taken as the smallest root of $x x'$, say δ_n , and V becomes

$$V = P_{\cdot n} \quad (40)$$

Perhaps the simplest way to solve for δ_n and $P_{\cdot n}$ is by writing

$$((\text{tr } x x') I - x x') V = V (\text{tr } (x x') - \delta_n) \quad (41)$$

and iterating equation 41 by some adaptation of the Hotelling method.

Suppose now

$$n > 2(m-1) \quad (42)$$

Then we may write

$$\begin{bmatrix} x_{1(j)}' \\ x_{2(j)}' \end{bmatrix} V = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (43)$$

and by the preceding methods we can write (44)

$$V = \left[I - \begin{bmatrix} x_{1(j)} & x_{2(j)} \end{bmatrix} \begin{bmatrix} x_{1(j)}' & x_{1(j)} & x_{1(j)}' & x_{2(j)} \\ x_{2(j)}' & x_{1(j)} & x_{2(j)}' & x_{2(j)} \end{bmatrix}^{-1} \begin{bmatrix} x_{1(j)}' \\ x_{2(j)}' \end{bmatrix} \right]^{-1} 1$$

If $n \leq 2(m-1)$, we let

$$\begin{bmatrix} x_1(j)' \\ x_2(j)' \end{bmatrix} V = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad (45)$$

$$b_1' b_1 + b_2' b_2 = \phi = \min. \quad (46)$$

$$(x_1(j), x_2(j)) = P \delta (Q_1', Q_2') \quad (47)$$

$$(P \delta^2 P' - \phi I) V = 0 \quad (48)$$

where P is $n \times m$

$$\phi = \delta_n^2 \quad (49)$$

and

$$V = P \cdot n \quad (50)$$

Consider now

$$C_1 = u_1' (I - \frac{1}{n} l l') u_1 / n \quad (51)$$

$$G_1 = u_1 (I - \frac{1}{n} l l') u_1' / m \quad (52)$$

$$C_2 = u_2' (I - \frac{1}{n} l l') u_2 / n \quad (53)$$

$$G_2 = u_2 (I - \frac{1}{n} l l') u_2' / m \quad (54)$$

Let

$$Z_1 = D x_1 \quad (55)$$

$$Z_2 = D x_2 \quad (56)$$

From equations 7, 8, 55, and 56

$$u_1 = Z_1 + B l' \quad (57)$$

$$u_2 = Z_2 + B l' \quad (58)$$

From equations 51, 57, 11, and 9 (59)

$$C_1 = (Z_1' Z_1 + Z_1' B l' + l B' Z_1 + l B' B l' - \frac{b_1 b_1'}{n}) / n$$

From equations 52, 57, 55, and 3

$$G_1 = Z_1 Z_1' / m \quad (60)$$

Similarly

$$C_2 = (Z_2' Z_2 + Z_2' B 1' + 1 B' Z_2 + 1 B' B 1' - \frac{b_2' b_2}{n}) / n \quad (61)$$

$$G_2 = Z_2 Z_2' / m \quad (62)$$

Let

$$\bar{\sigma}_{C1}^2 = \text{tr } C_1 / m \quad (63)$$

$$\bar{\sigma}_{G1}^2 = \text{tr } G_1 / n \quad (64)$$

$$\bar{\sigma}_{C2}^2 = \text{tr } C_2 / m \quad (65)$$

$$\bar{\sigma}_{G2}^2 = \text{tr } G_2 / n \quad (66)$$

From equations 55, 5, and 12

$$\text{tr } Z_1 Z_1' = m V' V \quad (67)$$

From equations 59, 4, 55, and 67

$$\bar{\sigma}_{C1}^2 = \frac{V' V}{n} + \frac{B' B}{n} - \frac{b_1' b_1}{n^2 m} \quad (68)$$

From equation 60

$$\bar{\sigma}_{G1}^2 = \frac{V' V}{n} \quad (69)$$

Similarly

$$\bar{\sigma}_{C2}^2 = \frac{V' V}{n} + \frac{B' B}{n} - \frac{b_2' b_2}{n^2 m} \quad (70)$$

$$\bar{\sigma}_{G2}^2 = \frac{V' V}{n} \quad (71)$$

Suppose we specify that

$$\bar{\sigma}_{C1}^2 - \bar{\sigma}_{G1}^2 = 0 \quad (72)$$

$$\bar{\sigma}_{C2}^2 - \bar{\sigma}_{G2}^2 = 0 \quad (73)$$

From equations 68, 69, and 72

$$B' B = \frac{b_1' b_1}{n m} \quad (74)$$

and likewise

$$B' B = \frac{b_2' b_2}{n m} \quad (75)$$

From equations 74 and 75 we have

$$b_1' b_1 = b_2' b_2 \quad (76)$$

A sufficient condition that equation 76 holds is

$$b_1 = b_2 \quad (77)$$

A necessary condition that equation 76 holds is

$$b_1 = b_2 h \quad (78)$$

where

$$h' h = h h' = I \quad (79)$$

The solution for V which satisfies equation 77 is given by equation 39.

Consider now

$$C_{12} = u_1' (I - \frac{1 1'}{n}) u_2 / n \quad (80)$$

$$G_{12} = u_1 (I - \frac{1 1'}{n}) u_2' / m \quad (81)$$

By methods analagous to those for obtaining equations 59 through 62 it can be shown, from equations 80 and 81, that

$$\text{tr } C_{12} / m = V' D_R V / n + \frac{B' B}{n} - \frac{b_1' b_2}{n^2 m} \quad (82)$$

$$\text{tr } G_{12} / n = V' D_R V / n \quad (83)$$

or from equations 74, 75, and 77 in equation 82

$$\text{tr } C_{12} / m = V' D_R V / n \quad (84)$$

hence

$$\text{tr } C_{12} / m = \text{tr } G_{12} / n \quad (85)$$

We see therefore that, if equation 77 holds (which it will if $n < m - 1$)

and B is chosen to satisfy equations 74 and 75, then:

1. The average row and column variances of u_1 and u_2 are all equal.
2. The average row covariance of u_1 and u_2 is equal to their average column covariance.

We next consider a determination of B . So far, equations 11, 74, and 75 are the only restrictions we have on it. Suppose we attempt a solution such that all the column variances of both matrices are as nearly equal as possible with the constraints in equations 11, 74, and 75.

Let

$$Y_1 = (I - \frac{1 \cdot 1'}{n}) Z_1 \quad (86)$$

$$Y_2 = (I - \frac{1 \cdot 1'}{n}) Z_2 \quad (87)$$

From equations 7, 51, 56, and 86

$$C_1 = (Y_1' Y_1 + Y_1' B 1' + 1 B' Y_1 + B' B \frac{1 \cdot 1'}{n}) / n \quad (88)$$

Similarly

$$C_2 = (Y_2' Y_2 + Y_2' B 1' + 1 B' Y_2 + B' B \frac{1 \cdot 1'}{n}) / n \quad (89)$$

Let

$$D_{C_1} 1 - k I = \epsilon_1 \quad (90)$$

$$D_{C_2} 1 - k I = \epsilon_2 \quad (91)$$

From equations 88, 89, and 90

$$Y_1^{(2)'} 1 + 2 Y_1' B + \frac{B' B}{n} 1 - k 1 = \epsilon_1 \quad (92)$$

Let

$$V_1 = Y_1^{(2)'} 1 / 2 \quad (93)$$

$$V_2 = Y_2^{(2)'} 1 / 2 \quad (94)$$

$$(\frac{B' B}{n} - k) = \gamma \quad (95)$$

From equations 92 through 95

$$V_1 + 2 Y_1' B + \gamma 1 = \epsilon_1 \quad (96)$$

Let

$$\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} = \begin{bmatrix} Y_1' \\ Y_2' \end{bmatrix} B + \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + \gamma \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (97)$$

Then

$$\begin{aligned} \phi = \epsilon_1' \epsilon_1 + \epsilon_2' \epsilon_2 = & B' (Y_1 Y_1' + Y_2 Y_2') B + 2 B' (Y_1 V_1 + Y_2 V_2) + \\ & + V_1' V_1 + V_2' V_2 + 2 \gamma (V_1' 1 + V_2' 1) + \gamma^2 2 m \end{aligned} \quad (98)$$

Let

$$\theta = \phi + B' B \lambda \quad (99)$$

$$\frac{\partial \theta}{\partial \gamma} = 0 \quad (100)$$

From equations 98, 99, and 100

$$\gamma = \frac{-1' (V_1 + V_2)}{2 m} \quad (101)$$

From equations 97 and 101

$$\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} = \begin{bmatrix} Y_1' \\ Y_2' \end{bmatrix} B + \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1' (V_1 + V_2)}{2 m} \quad (102)$$

Let

$$v_1 = V_1 + \gamma 1 \quad (103)$$

$$v_2 = V_2 + \gamma 1 \quad (104)$$

From equations 102, 103, and 104

$$\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} = \begin{bmatrix} Y_1' \\ Y_2' \end{bmatrix} B + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad (105)$$

From equations 98, 99, and 105

$$\theta = B' (Y_1 Y_1' + Y_2 Y_2') B + 2 B' (Y_1 v_1 + Y_2 v_2) - B' B \lambda$$

From equations 100 and 106

$$(Y_1 Y_1' + Y_2 Y_2' - \lambda I) B = -(Y_1 v_1 + Y_2 v_2) \quad (107)$$

Let

$$Y_1 Y_1' + Y_2 Y_2' = P \delta^2 P' \quad (108)$$

$$w = -(Y_1 v_1 + Y_2 v_2) \quad (109)$$

From equations 107 through 109

$$(P \delta^2 P' - \lambda I) B = w \quad (110)$$

From equation 110

$$(\delta^2 - \lambda I) P' B = P' w \quad (111)$$

From equation 111

$$P' B = (\delta^2 - \lambda I)^{-1} P' w \quad (112)$$

From equation 112

$$B = P (\delta^2 - \lambda I)^{-1} P' w \quad (113)$$

From equation 113

$$B' B - w' P (\delta^2 - \lambda I)^{-2} P' w = 0 \quad (114)$$

From equation 113 we find the smallest root. This can be proved to lie between 0 and δ_s^2 (the smallest δ_j^2). It can also be proved that only one such root lies in this interval.

As a matter of computational convenience, if n is much larger than $2m$, we can write the basic structure form

$$(Y_1, Y_2) = P \delta (Q_1', Q_2') \quad (115)$$

so that

$$\begin{bmatrix} Y_1' \\ Y_2' \end{bmatrix} \begin{bmatrix} Y_1, Y_2 \end{bmatrix} = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} \delta^2 \begin{bmatrix} Q_1', Q_2' \end{bmatrix} \quad (116)$$

The right side of equation 116 may be solved for δ^2 and the Q's by conveniently available eigen programs. Then for P we have

$$P = \begin{bmatrix} Y_1 & Y_2 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} \delta^{-1} \quad (117)$$

The smallest root of equation 114 may be readily solved for. Substituting this and P from equation 117, we can solve for B in equation 113.

APPENDIX G
MINIMUM VARIANCE TRANSFORMATION VECTORS
FOR PAIRED MATRICES

APPENDIX G

MINIMUM VARIANCE TRANSFORMATION VECTORS FOR PAIRED MATRICES

Suppose we have given two matrices X_1 and X_2 on the same n entities and m attributes. They may have been obtained at different times under different experimental conditions and different quantification strategies. We attempt linear row transformations on the two matrices so that in some defined sense they will be optimally similar. Let us require that:

1. Row mean and variance vectors for the two be equal;
2. Column means and variance vectors be as nearly equal as possible;
3. The variances of the transformation vectors be as small as possible.

1. Notation and Definition

Let X_1 and X_2 be $n \times m$ matrices.

$$x_1 = A_1 (X_1 - G_1 1 1') \quad (1.1)$$

$$1' x_1 1 = 0 \quad (1.2)$$

$$1' x_1^{(2)} 1 = n m \quad (1.3)$$

$$x_2 = A_2 (X_2 - G_2 1 1') \quad (1.4)$$

$$1' x_2 1 = 0 \quad (1.5)$$

$$1' x_2^{(2)} 1 = n m \quad (1.6)$$

$$z_1 = D_1^{-1} (x_1 + B_1 1') \quad (1.7)$$

$$z_1 1 = 0 \quad (1.8)$$

$$D_1 z_1 z_1' = m I \quad (1.9)$$

$$z_2 = D_2^{-1} (x_2 + B_2 1') \quad (1.10)$$

$$z_2 l = 0 \quad (1.11)$$

$$D_{z_2 z_2'} = m I \quad (1.12)$$

$$y_1 = D z_1 \quad (1.13)$$

$$y_2 = D z_2 \quad (1.14)$$

$$U_1 = y_1 + B l' \quad (1.15)$$

$$U_2 = y_2 + B l' \quad (1.16)$$

$$V = D l \quad (1.17)$$

$$B' l = 0 \quad (1.18)$$

$$l' U_1^{(2)} l = n m \quad (1.19)$$

$$l' U_2^{(2)} l = n m \quad (1.20)$$

$$u_1 = (I - \frac{l l'}{n}) U_1 \quad (1.21)$$

$$u_2 = (I - \frac{l l'}{n}) U_2 \quad (1.22)$$

$$y_1 = (I - \frac{l l'}{n}) Y_1 \quad (1.23)$$

$$y_2 = (I - \frac{l l'}{n}) Y_2 \quad (1.24)$$

From (1.15), (1.18), (1.21), and (1.23)

$$u_1 = y_1 + B l' \quad (1.25)$$

From (1.16), (1.18), (1.22), and (1.24)

$$u_2 = y_2 + B l' \quad (1.26)$$

From (1.13) and (1.15)

$$U_1^{(2)} = Y_1^{(2)} + 2 D_B Y_1 + B^{(2)} l' \quad (1.27)$$

From (1.14) and (1.16)

$$U_2^{(2)} = Y_2^{(2)} + 2 D_B Y_2 + B^{(2)} 1' \quad (1.28)$$

From (1.13) and (1.27)

$$U_1^{(2)} = D^2 z_1^{(2)} + 2 D D_B z_1 + B^{(2)} 1' \quad (1.29)$$

From (1.14) and (1.28)

$$U_2^{(2)} = D^2 z_2^{(2)} + 2 D D_B z_2 + B^{(2)} 1' \quad (1.30)$$

From (1.8), (1.9), (1.17), (1.19), and (1.29)

$$n m = m V' V + m B' B \quad (1.31)$$

From (1.11), (1.12), (1.17), (1.20), and (1.30)

$$n m = m V' V + m B' B \quad (1.32)$$

From (1.31) or (1.32)

$$1 = \frac{V' V}{n} + \frac{B' B}{n} \quad (1.33)$$

From (1.13), (1.15), (1.17), and (1.18)

$$U_1' 1 = z_1' V \quad (1.34)$$

From (1.14), (1.16), (1.17), and (1.18)

$$U_2' 1 = z_2' V \quad (1.35)$$

From (1.23)

$$u_1^{(2)} 1 = y_1^{(2)} 1 + 2 y_1' B + B' B 1 \quad (1.36)$$

From (1.24)

$$u_2^{(2)} 1 = y_2^{(2)} 1 + 2 y_2' B + B' B 1 \quad (1.37)$$

2. Constraints on the Transformed Matrices

Let

$$U_1' 1 - U_2' 1 = e \quad (2.1)$$

$$u_1^{(2)}, 1 - u_2^{(2)}, 1 = \epsilon \quad (2.2)$$

$$z_T = z_1 - z_2 \quad (2.3)$$

$$y_T = y_1 - y_2 \quad (2.4)$$

$$w_T = y_1^{(2)} - y_2^{(2)} \quad (2.5)$$

From (1.34), (1.35), and (2.3)

$$\epsilon_T = z_T' V \quad (2.6)$$

From (1.36), (1.37), (2.2), (2.4), and (2.5)

$$\epsilon_T = w_T' 1 + 2 y_T' B \quad (2.7)$$

Let

$$W_T = -w_T' 1 / 2 \quad (2.8)$$

$$E_T = \epsilon_T / 2 \quad (2.9)$$

From (2.7), (2.8), and (2.9)

$$y_T' B - W_T = E_T \quad (2.10)$$

We note now from (1.8), (1.11), (1.13), (1.14), and (2.4)

$$z_T' 1 = 0 \quad (2.11)$$

$$y_T' 1 = 0 \quad (2.12)$$

and from (1.19), (1.20), (2.5), and (2.8) that

$$w_T' 1 = 0. \quad (2.13)$$

We let

$$e_T = \begin{bmatrix} e \\ e_m \end{bmatrix} \quad (2.14)$$

$$z_T = \begin{bmatrix} z & z_m \end{bmatrix} \quad (2.15)$$

$$E_T = \begin{bmatrix} E \\ E_m \end{bmatrix} \quad (2.16)$$

$$y_T = \begin{bmatrix} y & y_m \end{bmatrix} \quad (2.17)$$

$$W_T = \begin{bmatrix} W \\ W_m \end{bmatrix} \quad (2.18)$$

Now, because of (2.11) through (2.18), we may write (2.6) and (2.10), respectively,

$$e = z' V \quad (2.19)$$

$$E = y' B - W \quad (2.20)$$

We shall now consider three special cases: (1) $n > m - 1$, (2) $n = m - 1$, and (3) $n < m - 1$.

3. Average Row and Column Variances

Consider now the average row variance of U_1 . This would be defined by

$$\sigma_{1R}^2 = \text{tr} (U_1 (I - \frac{11'}{m}) U_1') / n m \quad (3.1)$$

The average column variance of U_1 is given by

$$\sigma_{1C}^2 = \text{tr} (U_1' (I - \frac{11'}{n}) U_1) / n m \quad (3.2)$$

From (1.19) and (3.1)

$$\sigma_{1R}^2 = (n m - \frac{1' U_1' U_1 1}{m}) / n m \quad (3.3)$$

From (1.19) and (3.2)

$$\sigma_{1C}^2 = (n m - \frac{1' U_1 U_1' 1}{n}) / n m \quad (3.4)$$

From (1.8), (1.13), and (1.15)

$$U_1 = m B \quad (3.5)$$

From (1.13), (1.15), (1.17), and (1.18)

$$U_1 = V' z_1 \quad (3.6)$$

From (3.3) and (3.5)

$$\overline{\sigma_R^2} = 1 - \frac{B' B}{n} \quad (3.7)$$

From (3.4) and (3.6)

$$\overline{\sigma_C^2} = 1 - \frac{V' z_1 z_1' V}{n^2 m} \quad (3.8)$$

Similarly, we can show for U_2 that

$$\overline{\sigma_R^2} = 1 - \frac{B' B}{n} \quad (3.9)$$

and

$$\overline{\sigma_C^2} = 1 - \frac{V' z_2 z_2' V}{n^2 m} \quad (3.10)$$

From (1.33) and (3.7), and (3.9), respectively,

$$\overline{\sigma_R^2} = \frac{V' V}{n} \quad (3.11)$$

$$\overline{\sigma_R^2} = \frac{V' V}{n} \quad (3.12)$$

From (1.33) and (3.8), and (3.10), respectively,

$$\overline{\sigma_C^2} = \frac{V' V}{n} + \frac{B' B}{n} - \frac{V' z_1 z_1' V}{n^2 m} \quad (3.13)$$

$$\overline{\sigma_C^2} = \frac{V' V}{n} + \frac{B' B}{n} - \frac{V' z_2 z_2' V}{n^2 m} \quad (3.14)$$

From (3.11) and (3.12) we see that

$$\overline{\sigma_R^2} = \overline{\sigma_R^2} \quad (3.15)$$

Let us also require that

$$\overline{\sigma_C^2} = \overline{\sigma_C^2} \quad (3.16)$$

From (3.13), (3.14), and (3.16) we have

$$V' z_1 z_1' V = V' z_2 z_2' V \quad (3.17)$$

Let us also require that in (2.19)

$$e' e = \min. \quad (3.18)$$

We let

$$S_1 = z z' \quad (3.19)$$

$$S_2 = z_1 z_1' - z_2 z_2' \quad (3.20)$$

From (3.17) and (3.20)

$$V' S_2 V = 0 \quad (3.21)$$

Also, we let

$$V' V = k \quad (3.22)$$

and write

$$\psi = \text{tr} (e' e - V' S_2 V \lambda - V' V g) \quad (3.23)$$

From (2.19), (3.19), and (3.23)

$$\psi = V' (S_1 - S_2 \lambda - g I) V \quad (3.24)$$

Let

$$\frac{\partial \psi}{\partial V'} = 0 \quad (3.25)$$

From (3.24) and (3.25)

$$(S_1 - S_2 \lambda - g I) V = 0 \quad (3.26)$$

From (3.26)

$$V' S_2 (S_1 - S_2 \lambda - g I) V = 0 \quad (3.27)$$

From (3.21) and (3.27)

$$V' S_2 S_1 V - V' S_2^2 V \lambda = 0 \quad (3.28)$$

or

$$\lambda = \frac{V' S_2 S_1 V}{V' S_2^2 V} \quad (3.29)$$

In general, from (3.22), (3.26), and (3.29) we could solve iteratively for the smallest root g of (3.26) and the corresponding V .

4. The Non-Negative Row Scaling Vector

First, let us consider equation (2.19):

$$e = z' V$$

We wish to impose the restriction that

$$V \geq 0 \quad (4.1)$$

for all V_i . A necessary and sufficient condition for (4.1) is that D in (1.13) and (1.14) can be expressed by

$$D = d^2 \quad (4.2)$$

where d is real. We may write (2.19), because of (1.17) and (4.2),

$$e = z' d^2 1 \quad (4.3)$$

and (3.22) becomes

$$k = 1' d^4 1 \quad (4.4)$$

We let

$$v = d l \quad (4.5)$$

Then (4.3) and (4.4) become, respectively,

$$e = z' d v \quad (4.6)$$

$$k = v' d^2 v \quad (4.7)$$

Let

$$\phi = \text{tr } e' e \quad (4.8)$$

Now instead of imposing constraint (3.21), we may wish to minimize (4.8) with constraint (4.7). A sufficient condition for (3.17) to be satisfied identically is that

$$z_1' v = z_2' v \quad (4.9)$$

If (4.9) is satisfied identically, then ϕ in (4.8) vanishes.

We set up the function

$$\psi = \phi - \lambda v' d^2 v \quad (4.10)$$

and require that

$$\frac{\partial \psi}{\partial v'} = 0 \quad (4.11)$$

From (4.4) and (4.10)

$$\psi = v' d z z' d v - v' d^2 v \lambda \quad (4.12)$$

From (4.11) and (4.12)

$$(d z z' d - \lambda d^2) v = 0 \quad (4.13)$$

5. The Row Scaling Vector for $n = m$

Now if $n > m - 1$ and z is basic, we can always find a V which will make e in (2.19) vanish. In particular, suppose $n = m$ and the basic structure of z is

$$z = P \Delta Q' \quad (5.1)$$

Then by definition Q is square and P is of order $m \times (m - 1)$. One and only one V will make e vanish in (2.19) and it may readily be found as follows:

V will be the vector orthogonal to P , i.e.,

$$P' V = 0 \quad (5.2)$$

Since P lacks being square by only a single vector, we can write the unique orthonormal matrix

$$\rho = (P, p) \quad (5.3)$$

where p is the unit length vector orthogonal to P and hence proportional to V . By definition then

$$\rho \rho' = I \quad (5.4)$$

From (5.3) and (5.4)

$$P P' + p p' = I \quad (5.5)$$

From (5.5)

$$P P' 1 + p p' 1 = 1 \quad (5.6)$$

or

$$p = (1 - P (P' 1)) k \quad (5.7)$$

where k is a normalizing constant for p . Then we have

$$V = (1 - P P' 1) g \quad (5.8)$$

where g is a constant such that $V' V$ is some value to be determined.

Suppose, however, that $n - m = t$ where t is greater than 1. Then in (5.7) p would be of width S and we could write

$$V = p \nu \quad (5.9)$$

where v is any vector scaled so that $V' V$ is the prespecified value, and clearly e in (2.19) would vanish.

6. The Minimum Variance Row Scaling Vector

Now if p is of width greater than 1, we have degrees of freedom left for determining V and additional constraints are required. A sensible and interesting rationale is that the variation in the elements of V shall be as small as possible while at the same time making e vanish in (2.19). This rationale is simpler to achieve analytically than that all elements in V shall be non-negative. We proceed as follows: Let

$$e = V - k \mathbf{1} \quad (6.1)$$

where k is a constant to be determined. Let

$$\psi = e' e - 2 V' z \lambda \quad (6.2)$$

where λ is a vector of Lagrangian multipliers.

From (6.1) and (6.2)

$$\psi = V' V - 2 V' \mathbf{1} k + k^2 \mathbf{1}' \mathbf{1} + 2 V' z \lambda \quad (6.3)$$

Take the symbolic derivative of ψ with respect to V' and equate to zero, thus,

$$\frac{\partial \psi}{\partial V'} = 0. \quad (6.4)$$

From (6.3) and (6.4)

$$V - \mathbf{1} k + z \lambda = 0 \quad (6.5)$$

Premultiplying (6.5) by z' and setting

$$z' V = 0 \quad (6.6)$$

we get

$$z' z \lambda = z' \mathbf{1} k. \quad (6.7)$$

From (6.7)

$$\lambda = (z' z)^{-1} z' l k \quad (6.8)$$

Substituting (6.8) in (6.5)

$$V - l k + z (z' z)^{-1} z' l k = 0 \quad (6.9)$$

From (6.9)

$$V = (1 - z (z' z)^{-1} z' l) k \quad (6.10)$$

Then k in (6.10) can be determined so that $V' V$ is the pre-specified value.

7. The Scaling Vector for $n < m$

If $n \leq m - 1$, then e in (2.19) cannot be zero. We then consider the minimization of $e' e$ with a constraint on $V' V$. We let

$$\psi = e' e - \lambda V' V \quad (7.1)$$

From (2.19) and (5.1)

$$\psi = V' P \Delta^2 P' V - V' V \lambda \quad (7.2)$$

Let

$$\frac{\partial \psi}{\partial V'} = 0 \quad (7.3)$$

From (7.2) and (7.3)

$$(P \Delta^2 P' - \lambda I) V = 0 \quad (7.4)$$

Let Δ_n^2 and $P_{.n}$ be the smallest element of Δ^2 and the corresponding vector of P . Then for $e' e$ a minimum it is well known that we must have

$$V = P_{.n} k \quad (7.5)$$

where k is determined to yield the prespecified value of $V' V$.

8. The Additive Vector for $n > m - 1$

Next we return to equation (2.20) and consider the determination of the additive vector B . This equation is

$$E = y' B - W$$

where y and W have been previously defined.

First we shall consider the case where $n > m - 1$. Then assuming y basic, we can always write

$$y' B = W \quad (8.1)$$

so that E in (2.20) vanishes.

The general solution for B in (8.1) is

$$B = (I - y (y' y)^{-1} y') b + y (y' y)^{-1} W \quad (8.2)$$

where

$$b' 1 = 0 \quad (8.3)$$

and

$$y' b \neq 0.$$

Next, from (8.2) we can prove that

$$B' B = b' (I - y (y' y)^{-1} y') b + W' (y' y)^{-1} W \quad (8.5)$$

Now b is arbitrary except for (8.3) and (8.4). It is also clear from (8.5) that the minimum value $B' B$ can take is

$$B' B = W' (y' y)^{-1} W \quad (8.6)$$

This is the case when b in (8.2) is zero. The maximum value is unrestricted except for (1.33). We may select b in (8.2) at will to satisfy any specified value of $B' B$ in (8.5) so long as (8.3) and (8.4) are satisfied.

9. The Average Column Covariance

Let us see now how the average of column covariance of U_1

with U_2 is affected by the determination of B . This average can be shown to be

$$\overline{C}_c = \text{tr } U_1' (I - \frac{1}{n} 11') U_2 / m n \quad (9.1)$$

From (1.15), (1.16), and (9.1)

$$\overline{C}_c = \text{tr } (Y_1' + 1 B') (I - \frac{1}{n} 11') (Y_2 + B 1') / m n \quad (9.2)$$

From (1.18) and (9.2)

$$\overline{C}_c = \text{tr } ((Y_1' + 1 B') (Y_2 + B 1') - Y_1' \frac{1}{n} 11' Y_2) / m n \quad (9.3)$$

From (1.8), (1.11), (1.13), (1.14), and (9.3)

$$\overline{C}_c = (\text{tr } (Y_1' Y_2') + m B' B - \frac{V' z_1' z_2 V}{n}) / n m \quad (9.4)$$

As we have seen, if $n > m - 1$, V can be determined so that $V' z_1 = V' z_2$ or, of $z = z_1 - z_2$, then $z' V = 0$. (See (6.6)). Suppose we let

$$M_c = \frac{z_1' V}{n} = \frac{z_2' V}{n} \quad (9.5)$$

From (1.9), (1.11), and (9.5)

$$1' M_c = 0 \quad (9.6)$$

Let

$$D_r = D_{z_1 z_2'} / m \quad (9.6)$$

Then D_r is a diagonal of the row correlations of x_1 with x_2 .

From (1.13), (1.14), (1.17), and (9.7)

$$\text{tr } Y_1' Y_2' = m V' D_r V \quad (9.8)$$

From (9.4), (9.5) and (9.8)

$$\overline{C}_c = \frac{V' D_r V}{n} + \frac{B' B}{n} - \frac{M_c' M_c}{m} \quad (9.9)$$

From (3.13), (3.14), and (9.5)

$$\overline{\sigma_c^2} = \frac{V' V}{n} + \frac{B' B}{n} - \frac{M_c' M_c}{m} \quad (9.10)$$

The average row covariance for U_1 with U_2 is independent of B and is simply

$$\overline{C_R} = V' D_{z_1 z_2} V / n m \quad (9.11)$$

From (9.7) and (9.11)

$$\overline{C_R} = \frac{V' D_r V}{n} \quad (9.12)$$

Suppose now we write from (3.11) and (3.12)

$$\overline{\sigma_R^2} = \frac{V' V}{n} \quad (9.13)$$

and require that

$$\overline{\sigma_c^2} = \overline{\sigma_R^2} \quad (9.14)$$

From (9.10), (9.13), and (9.14)

$$\frac{B' B}{n} = \frac{M_c' M_c}{m} \quad (9.15)$$

or the variance of B is equal to the variance of the column means of U_1 and of U_2 . From (9.9), (9.14), and (9.15)

$$\overline{C_c} = \frac{V' D_r V}{n} \quad (9.16)$$

and from (9.13) and (9.16)

$$\overline{C_R} = \overline{C_c} \quad (9.17)$$

Therefore we see that, if B is determined so that the average row variance is equal to the average column variance, then:

1. The variance of row means for the U 's is equal to their average column variance.

2. The average covariance of rows between U_1 and U_2 is equal to the average covariance of columns between the two matrices.

We also see from (9.9) that the higher the variance of B or the column means of the U 's, the larger will be the average column covariance.

10. The Additive Vector for $n < m$

If $n \leq m - 1$, then E in (2.20) cannot in general vanish and we write

$$\phi = E' E \quad (10.1)$$

and the constraining condition

$$B' B = g^2. \quad (10.2)$$

We let

$$\psi = \phi + B' B \lambda \quad (10.3)$$

From (2.20), (10.1), and (10.3)

$$\psi = B' y y' B - 2 B' y W + W' W - B' B \lambda \quad (10.4)$$

We let

$$\frac{\partial \psi}{\partial B'} = 0 \quad (10.5)$$

From (10.4) and (10.5)

$$0 = (y y' - \lambda I) B - y W \quad (10.6)$$

Let the basic structure of y be

$$y = P_y \Delta_y Q_y' \quad (10.7)$$

From (10.6) and (10.7)

$$(P_y \Delta_y^2 P_y' - \lambda I) B - P_y \Delta_y Q_y' W \quad (10.8)$$

From (10.8)

$$(\Delta_y^2 - \lambda I) P_y' B - \Delta_y Q_y' W = 0 \quad (10.9)$$

From (10.9)

$$P_y' B = (\Delta_y^2 - \lambda I)^{-1} \Delta_y Q_y' W \quad (10.10)$$

Now because of (1.23) and (1.24)

$$P_y' 1 = 0 \quad (10.11)$$

so that the order of P_y is $n \times (n - 1)$. Let

$$\mu = \Delta_y Q_y' W \quad (10.12)$$

From (10.10)

$$B = P_y (\Delta_y^2 - \lambda I)^{-1} \mu \quad (10.13)$$

From (10.2) and (10.12)

$$g^2 - \mu' (\Delta_y^2 - \lambda I)^{-2} \mu \quad (10.14)$$

The smallest root λ of (10.14) when substituted in (10.12) is known to give the smallest ϕ in (10.1) for a specified g^2 in (10.2).

In particular, if we put no constraint on $B' B$, we let $\lambda = 0$, and get from (10.12) and (10.13)

$$B = P_y \Delta_y^{-1} Q_y' W \quad (10.15)$$

which is the general inverse solution of (8.1). From (10.15) then

$$B' B = W' Q_y \Delta_y^{-\frac{1}{2}} Q_y' W. \quad (10.16)$$

APPENDIX H

MAXIMUM CORRELATION TRANSFORMATION VECTORS FOR PAIRED MATRICES WITH EQUAL COLUMN MEANS AND VARIANCES

APPENDIX H

MAXIMUM CORRELATION TRANSFORMATION VECTORS FOR PAIRED MATRICES WITH EQUAL COLUMN MEANS AND VARIANCES

1. The Model, the Notation, and the Restrictions

Suppose we have given two matrices z_1 and z_2 for the same n entities and m attributes with n greater than m . We assume that z_1 and z_2 have been obtained from two other matrices, x_1 and x_2 , by a standardization of the rows of each x matrix. We assume further that by identical row transformations of the two matrices z_1 and z_2 we get two new matrices U_1 and U_2 . The row scaling parameters consist of a scaling vector D and an additive vector B . Since both the z matrices have been normalized by rows and the row transformations from z to U are the same for both matrices, the resulting U matrices will have the same row means and variances. Furthermore, the correlations of corresponding rows of U_1 and U_2 will be the same as for z_1 and z_2 .

We shall in this model also impose constraints on the D and B parameters so that the means and variances of the columns of U_1 and U_2 will be the same. But since n is greater than m , we shall have a surplus of degrees of freedom. We restrict B further so that $B' B$ is a minimum. With these constraints on the parameters of the z to U transformations we now wish to determine D and B so that the correlation between all the corresponding elements of the U_1 and U_2 matrices shall be a maximum.

Let

$$z_1 \cdot 1 = 0 \quad (1.1)$$

$$z_1^{(2)} \cdot 1 = m \cdot 1 \quad (1.2)$$

$$z_2 \cdot 1 = 0 \quad (1.3)$$

$$z_2^{(2)} \cdot 1 = m \cdot 1 \quad (1.4)$$

$$(z_1 \cdot z_2) \cdot 1 = m \cdot D_R \cdot 1 \quad (1.5)$$

$$U_1 = D \cdot z_1 + B \cdot 1' \quad (1.6)$$

$$U_2 = D \cdot z_2 + B \cdot 1' \quad (1.7)$$

$$B' \cdot 1 = 0 \quad (1.8)$$

$$1' \cdot U_1 = 1' \cdot U_2 \quad (1.9)$$

$$1' \cdot U_1^{(2)} = 1' \cdot U_2^{(2)} \quad (1.10)$$

$$V = D \cdot 1 \quad (1.11)$$

2. The Additive Vector as a Function of the Scaling Vector

From (1.6), (1.7), (1.8), (1.9), and (1.11)

$$V' \cdot (z_1 - z_2) = 0 \quad (2.1)$$

From (1.6)

$$U_1^{(2)} = D^2 \cdot z_1^{(2)} + 2 \cdot D_B \cdot D \cdot z_1 + B^{(2)} \cdot 1' \quad (2.2)$$

From (1.7)

$$U_2^{(2)} = D^2 \cdot z_2^{(2)} + 2 \cdot D_B \cdot D \cdot z_2 + B^{(2)} \cdot 1' \quad (2.3)$$

From (1.10), (1.11), (2.2), and (2.3)

$$(z_1^{(2)} - z_2^{(2)})' \cdot D \cdot V + 2 \cdot (z_1 - z_2)' \cdot D \cdot B = 0 \quad (2.4)$$

Let

$$F_1 = z_1 - z_2 \quad (2.5)$$

$$F_2 = (z_1^{(2)} - z_2^{(2)}) / 2 \quad (2.6)$$

From (2.1), (2.5), and (2.6)

$$V' F_1 = 0 \quad (2.7)$$

From (2.4), (2.5), and (2.6)

$$F_2' D V + F_1' D B = 0 \quad (2.8)$$

The general inverse solution of B in (2.5) also minimizes $B'B$ and is

$$B = -(D F_1 (F_1' D^2 F_1)^{-1} F_2' D V) \quad (2.9)$$

3. The Attribute Variances and Covariances

Let

$$C_1 = U_1' (I - \frac{1 \cdot 1'}{n}) U_1 / n \quad (3.1)$$

$$C_2 = U_2' (I - \frac{1 \cdot 1'}{n}) U_2 / n \quad (3.2)$$

$$C_{12} = U_1' (I - \frac{1 \cdot 1'}{n}) U_2 / n \quad (3.3)$$

From (1.6) and (1.8)

$$C_1 = (z_1' D^2 z_1 + z_1' D B 1' + 1 B' D z_1 + B' B 1 1' - z_1' \frac{D 1 1' D}{n} z_1) / n \quad (3.4)$$

Let

$$\bar{C}_1 = (\text{tr } C_1) / m \quad (3.5)$$

From (1.1), (1.8), (1.11), (3.4), and (3.5)

$$\bar{C}_1 = (V' D_{z_1 z_1} V + m B' B - \frac{V' z_1 z_1' V}{n}) / n m \quad (3.6)$$

By analogy

$$\bar{C}_2 = (V' D_{z_2 z_2} V + m B' B - \frac{V' z_2 z_2' V}{n}) / n m \quad (3.7)$$

and

$$\overline{C}_{12} = (V' D_{z_1 z_2'} V + m B' B - \frac{V' z_1 z_2' V}{n}) / n m \quad (3.8)$$

From (1.2), (1.4), and (1.5), respectively, we have

$$D_{z_1 z_1'} = m I \quad (3.9)$$

$$D_{z_2 z_2'} = m I \quad (3.10)$$

$$D_{z_1 z_2'} = m D_R \quad (3.11)$$

Let

$$z_T = (z_1 + z_2) \quad (3.12)$$

From (2.1) and (3.12)

$$V' z_1 z_1' V = \frac{1}{4} V' z_T z_T' V \quad (3.13)$$

$$V' z_2 z_2' V = \frac{1}{4} V' z_T z_T' V \quad (3.14)$$

$$V' z_1 z_2' V = \frac{1}{4} V' z_T z_T' V \quad (3.15)$$

From (3.6), (3.9), and (3.13)

$$\overline{C}_1 = (V' V - \frac{V' z_T z_T' V}{4 n m} + B' B) / n \quad (3.16)$$

Let

$$F_3 = z_T / 2 \sqrt{n m} \quad (3.17)$$

From (2.9)

$$B' B = V' D F_2 (F_1' D^2 F_1)^{(-1)} F_2' D V \quad (3.18)$$

Let

$$Q_2 \Delta_2^2 Q_2' = F_1' D^2 F_1 \quad (3.19)$$

$$F_4 = D F_2 Q_2 \Delta_2^{-1} \quad (3.20)$$

From (3.10) through (3.20)

$$\overline{C}_1 = V' (I - F_3 F_3' + F_4 F_4') V / n \quad (3.21)$$

Because of (3.10)

$$\overline{C}_2 = \overline{C}_1 \quad (3.22)$$

Because of (3.11)

$$\overline{C}_{12} = V' (D_R - F_3 F_3' + F_4 F_4') V / n \quad (3.23)$$

4. The Optimizing Criterion Function

Let

$$S = F_3 F_3' - F_4 F_4' \quad (4.1)$$

From (3.21) through (4.1)

$$n \overline{C}_1 = V' (I - S) V \quad (4.2)$$

$$n \overline{C}_2 = V' (I - S) V \quad (4.3)$$

$$n \overline{C}_{12} = V' (D_R - S) V \quad (4.4)$$

Now let

$$G_1 = D_R - S \quad (4.5)$$

$$G_2 = I - S \quad (4.6)$$

$$f_1 = V' G_1 V \quad (4.7)$$

$$f_2 = V' G_2 V \quad (4.8)$$

$$f_3 = V' F_1 \lambda \quad (4.9)$$

where λ is a Lagrangian vector.

$$\phi = \frac{f_1}{f_2} \quad (4.10)$$

$$\psi = \phi - f_3 \quad (4.11)$$

From (4.10) and (4.11)

$$d \psi = \frac{d f_1}{f_1} - \frac{f_1}{f_2^2} d f_2 - d f_3 \quad (4.12)$$

From (4.10) and (4.12)

$$\frac{1}{f_2} d \psi = d f_1 - f_2 d f_3 - \emptyset d f_2 \quad (4.13)$$

From (4.13)

$$\frac{1}{f_2} \frac{\partial \psi}{\partial V^i} = \frac{\partial f_1}{\partial V^i} - f_2 \frac{\partial f_3}{\partial V^i} - \emptyset \frac{\partial f_2}{\partial V^i} \quad (4.14)$$

5. The Solution for the Scaling Vector

From (4.7) for a fixed D in F_4

$$\frac{\partial f_1}{\partial V^i} = 2 G_1 V \quad (5.1)$$

From (4.8) for a fixed D in F_4

$$\frac{\partial f_2}{\partial V^i} = 2 G_2 V \quad (5.2)$$

From (4.9)

$$\frac{\partial f_3}{\partial V^i} = F_1 \quad (5.3)$$

Let

$$\frac{\partial \psi}{\partial V^i} = 0 \quad (5.4)$$

$$\gamma = 2 f_2 \lambda \quad (5.5)$$

From (4.14) through (5.5)

$$G_1 V - F_1 \gamma - G_2 V \emptyset = 0 \quad (5.6)$$

From (4.5), (4.6), and (5.6)

$$(D_R - S) V - F_1 \gamma - (I - S) V \emptyset = 0 \quad (5.7)$$

From (5.7)

$$(D_R - (1 - \emptyset) S) V - F_1 \gamma = \emptyset V \quad (5.8)$$

From (2.7) and (5.8)

$$F_1' (D_R - (1 - \emptyset) S) V - F_1' F_1 \gamma = 0 \quad (5.9)$$

From (5.9)

$$\gamma = (F_1' F_1)^{(-1)} F_1' (D_R - (1 - \emptyset) S) V \quad (5.10)$$

Let

$$Q_1 \Delta_1^2 Q_1' = F_1' F_1 \quad (5.11)$$

$$F_5 = F_1 Q_1 \Delta_1^{-1} \quad (5.12)$$

$$1 - \emptyset = g \quad (5.13)$$

From (4.1), (5.8), (5.11), (5.12), and (5.13)

$$(I - F_5 F_5') (D_R - g (F_3 F_3' - F_4 F_4')) V = V \emptyset \quad (5.14)$$

6. The Computational Equations

$$V' V = n \quad (6.1)$$

$$W_1 = F_4' V g \quad (6.2)$$

$$W_2 = F_3' V g \quad (6.3)$$

$$V = D_R V - F_3 W_2 + F_4 W_1 \quad (6.4)$$

$$W_5 = F_5' V \quad (6.5)$$

$$V = V - F_5 W_5 \quad (6.6)$$

$$\emptyset = \frac{V' V}{n} \quad (6.7)$$

$$V = V / \sqrt{\emptyset} \quad (6.8)$$

$$g = 1 - \emptyset \quad (6.9)$$

First calculate

$$F_1 = z_1 - z_2 \quad (2.5)$$

$$F_2 = (z_1^{(2)} - z_2^{(2)}) / 2 \quad (2.6)$$

$$F_3 = (z_1 + z_2) / 2 \sqrt{n m} \quad (3.12) \text{ \& } (3.17)$$

$$Q_1 \Delta_1^2 Q_1' = F_1' F_1 \quad (5.11)$$

$$F_5 = F_1 Q_1 \Delta_1^{-1} \quad (5.12)$$

Then calculate iteratively

$$Q_2 \Delta_2^2 Q_2' = F_1' D^2 F_1 \quad (3.19)$$

$$F_4 = D F_2 Q_2 \Delta_2^{-1} \quad (3.20)$$

and equations (6.2) through (6.9).

For initial values of V and g let

$$V = 1 \quad (6.10)$$

$$g = 1 - 1' D_R 1 / n \quad (6.11)$$

When \emptyset stabilizes sufficiently

$$W_2 = Q_2 \Delta_2^{-1} F_4' V \quad (6.12)$$

$$B = -D F_1 W_2 \quad (6.13)$$

7. The Correlation of Corresponding U Elements

Let us now define $1^z.j$ as the j th column vector of z_1 , $2^z.j$ as the j th column vector of z_2 , $1^U.j$ as the j th column vector of U_1 , and $2^U.j$ as the j th column vector of U_2 .

We let

$$V_{1^U.j} \equiv \begin{bmatrix} 1^U.1 \\ \vdots \\ 1^U.m \end{bmatrix} \quad (7.1)$$

$$V_{2^U.j} \equiv \begin{bmatrix} 2^U.1 \\ \vdots \\ 2^U.m \end{bmatrix} \quad (7.2)$$

$$D_{1^Z.j} \equiv \begin{bmatrix} 1^Z.1j & \dots & 0 \\ \cdot & \cdot & \cdot \\ 0 & \dots & 1^Z.nj \end{bmatrix} \quad (7.3)$$

$$D_{2^Z.j} \equiv \begin{bmatrix} 2^Z.1j & \dots & 0 \\ \cdot & \cdot & \cdot \\ 0 & \dots & 2^Z.nj \end{bmatrix} \quad (7.4)$$

$$V_{1^{D_Z}.j} \equiv \begin{bmatrix} D_{1^Z.1} \\ \vdots \\ D_{1^Z.m} \end{bmatrix} \quad (7.5)$$

$$V_{2^{D_Z}.j} \equiv \begin{bmatrix} D_{2^Z.1} \\ \vdots \\ D_{2^Z.m} \end{bmatrix} \quad (7.6)$$

$$V_I \equiv \begin{bmatrix} I \\ \vdots \\ I \end{bmatrix}_m \quad (7.7)$$

$$D_{\beta} 1 = -D F_1 (F_1' D^2 F)^{-1} F_2' D 1 \quad (7.8)$$

From (1.11) and (2.9)

$$D_{\beta} V = B \quad (7.9)$$

From (1.6), (1.7), and (7.1) through (7.9) it can be proved that

$$V_{1^U.j} = (V_D \quad + V_I D_B) V \quad (7.10)$$

$$V_{2^U.j} = (V_D \quad + V_I D_B) V \quad (7.11)$$

From (1.1), (1.6), (7.8), and (7.10) it can be proved that

$$1' V_{1^U.j} = 0 \quad (7.12)$$

Similarly, it can be proved that

$$1' V_{2^U.j} = 0 \quad (7.13)$$

From (4.7), (4.8), and (4.9), respectively, it can be proved that

$$V_{1^U.j} \cdot V_{2^U.j} = f_1 \quad (7.14)$$

$$V_{1^U.j} \cdot V_{1^U.j} = f_2 \quad (7.15)$$

$$V_{2^U.j} \cdot V_{2^U.j} = f_2 \quad (7.16)$$

But from (7.12) through (7.16) we see that ϕ in (4.10) is precisely the correlation between the $V_{1^U.j}$ and $V_{2^U.j}$ vectors with the constraints (1.9) and (1.10) and the further constraint that $B' B$ is minimized.

8. The Estimate of the Unbiased Correlation

The solution for maximizing ψ in (4.11) involves n m observations and n parameters in V . But (1.9) and (1.10) impose 2 m constraints and the solution for B in (2.9) an additional constraint. Hence we have only $n - 2$ $m - 1$ free

parameters.

To estimate the unbiased ϕ in (4.10) we suggest

$$\phi_c^2 = \frac{\phi_c^2 n m - (n - 2 m - 1)}{n m - (n - 2 m - 1)} \quad (8.1)$$

This estimate should be compared with

$$\phi_e^2 = \left(\frac{l' D_R l}{n} \right)^2 \quad (8.2)$$

APPENDIX I
THE NEBEKER MODEL

APPENDIX I

THE NEBEKER MODEL

Nebeker has proposed*an interesting expectancy theory model for predicting performance, having given (1) the valences of performance levels and (2) a matrix of the conditional expectancies that a given effort level will be followed by a given performance level. The description which follows applies to a single person but it could readily be extended to a group of persons.

I. Definition of Notation

We define notation as follows:

a is the effort dimension

b is the performance dimension

n_a is the number of levels in dimension a

n_b is the number of levels in dimension b

L_{bj} is the value assigned to level j of dimension b

L_b is the n_b th order vector of the L_{bj}

P_{bj} is the proportion of time given to level j of dimension b

P_b is the n_b th order vector of the P_{bj}

P_{aibj} is the conditional expectancy that level i of dimension a will be followed by level j of dimension b

P_{ab} is the $n_a \times n_b$ matrix of the P_{aibj}

P_{ba} is the transpose of P_{ab}

V_{bj} is the valence of level j of dimension b

V_b is the n_b th order vector of the V_{bj}

P_{aibj} is the expectancy of the joint occurrence of level i of dimension a and level j of dimension b

* Personal communication.

p_{ab} is the $n_a \times n_b$ matrix of the $p_{a_i b_j}$

p_{ba} is the transpose of p_{ab}

p_{a_i} is the expectancy that level i of dimension a will occur

p_a is an n_a th order vector of the p_{a_i}

p_{b_j} is the expectancy that level j of dimension b will occur

p_b is the n_b th order vector of the p_{b_j}

\tilde{V}_{a_i} is the ^{estimated} valence of level i of dimension a

\tilde{V}_a is the n_a th order vector of the \tilde{V}_{a_i}

D_a is an n_a th order diagonal matrix of the p_{a_i}

II. The Assumptions

We assume that P_{ab} and V_b are given. By definition

$$p_a = p_{ab} l \quad (2.1)$$

$$p_b = p_{ba} l \quad (2.2)$$

$$P_{ab} = D_a^{-1} p_{ab} \quad (2.3)$$

$$l' p_a = 1 \quad (2.4)$$

$$l' p_b = 1 \quad (2.5)$$

$$P_{ab} l = 1 \quad (2.6)$$

If we let Y be the observed productivity and \tilde{Y} the estimated productivity, then

$$Y = P_b' L_b \quad (2.7)$$

$$\tilde{Y} = p_b' L_b \quad (2.8)$$

We wish to estimate p_b which is unknown. To do this we make two assumptions.

(1) The vector V_a of estimated effort level valences is the product of the

conditional expectancy matrix by the vector of performance level valences.

(2) The unknown vector p_a of performance level expectancies is proportional to the effort level valences. The first of these is in line with traditional expectancy theory modeling. The second seems intuitively reasonable except for a possible generalization of the relationship between the two vectors which we shall consider presently.

III. The Derivations

First we shall see how the p_b vector is estimated on the basis of the two assumptions. According to these two, we have, respectively,

$$V_a = P_{ab} V_b \quad (3.1)$$

$$p_a = c V_a \quad (3.2)$$

where c is the proportionality constant. From (2.3) we have

$$P_{ab} = D_a P_{ab} \quad (3.3)$$

From (2.2) and (3.3)

$$P_b = P_{ba} D_a 1 \quad (3.4)$$

From the definition of D_a

$$p_a = D_a 1 \quad (3.5)$$

From (3.4) and (3.5)

$$P_b = P_{ba} p_a \quad (3.6)$$

From (2.4) and (3.2)

$$1 = c 1' V_a \quad (3.7)$$

From (3.7)

$$c = 1 / 1' V_a \quad (3.8)$$

From (3.2), (3.6), and (3.8)

$$p_b = \frac{P_{ba} V_a}{1' V_a} \quad (3.9)$$

Premultiplying (3.9) by $1'$ we have

$$1' p_b = \frac{1' P_{ba} V_a}{1' V_a} \quad (3.10)$$

or from (2.6) and (3.10)

$$1' p_b = 1 \quad (3.11)$$

which is, as it must be, the same as equation (2.5).

IV. The Calculations

The sequence of calculations from the equations above is

$$V_a = P_{ab} V_b \quad (3.1)$$

$$c = 1 / 1' V_a \quad (3.8)$$

$$p_a = c V_a \quad (3.2)$$

$$p_b = P_{ba} p_a \quad (3.6)$$

To get \tilde{Y} , the estimated performance, we use equation (2.8).

Although we have some reluctance in emphasizing the "across-subjects vs. within-subjects analysis" concept, it is possible to do a within-subjects analysis by comparing the observed P_b and the estimated p_b . The correlation between the two could be computed.

V. Generalized Linear Relationship

Now instead of assuming the relationship given by (3.2), we could assume a generalized linear relationship between p_a and V_a , thus:

$$p_a = c_0 1 + c V_a \quad (5.1)$$

where c_0 is an additive constant and c is a scaling constant. From (2.4), (3.1), and (5.1)

$$1 = c_0 n_a + c l' v_a \quad (5.2)$$

From (5.2)

$$c_0 = \frac{1 - c l' v_a}{n_a} \quad (5.3)$$

From (5.1) and (5.3)

$$p_a = \frac{1 (1 - c l' v_a)}{n_a} + c v_a \quad (5.4)$$

From (5.4)

$$p_a = 1 / n_a + c (I - \frac{1 l'}{n_a}) v_a \quad (5.5)$$

But the expression in parentheses on the right of (5.5) is a deviating matrix for v_a . We may therefore write

$$v_a = (I - \frac{1 l'}{n_a}) v_a \quad (5.6)$$

so that

$$l' v_a = 0 \quad (5.7)$$

From (5.5) and (5.7)

$$p_a = 1 / n_a + c v_a \quad (5.8)$$

From (3.6) and (5.8)

$$p_b = \frac{p_{ba} l}{n_a} + c p_{ba} v_a \quad (5.9)$$

From (2.6), (5.7), and (5.9) we have

$$l' p_b = 1 \quad (5.10)$$

which again is, as it must be, the same as equation (2.5).

We can now solve for a single c for all persons which should give a more stable and useful solution for the parameter c .

We could calculate a least squares solution for c but again a better solution, analogous to (5.21), might be

$$c = \frac{\sigma_b^f}{\sigma_b^u} \quad (6.4)$$

The number of observations is $n_a \times n$ instead of only n_a when we consider all persons. We may therefore write, instead of (5.6), the general polynomial function

$$p_a = c_0 \cdot 1 + c_1 v_a + \dots + c_2 v_a^{(m)} \quad (6.5)$$

We impose the constraints of equation (2.5), whereupon we can write from (6.5)

$$p_a = \frac{1}{n_a} + (I - \frac{1 \cdot 1'}{n_a}) (c_1 v_a + \dots + c_2 v_a^{(m)}) \quad (6.6)$$

We let

$$v_{a_k} = (I - \frac{1 \cdot 1'}{n_a}) v_a^{(k)} \quad (6.7)$$

From (6.6) and (6.7)

$$p_a = \frac{1}{n_a} + c_1 v_{a_1} + \dots + c_2 v_{a_m} \quad (6.8)$$

We let

$$p_{ba} v_{a_k} = \mu_{b_k} \quad (6.9)$$

From (3.6), (5.11), (6.8), and (6.9)

$$p_b = v_b + c_1 \mu_{b_1} + \dots + c_2 \mu_{b_m} \quad (6.10)$$

Suppose now we redefine u_b in (5.12) as the matrix

$$u_b \equiv (u_{b_1}, \dots, u_{b_m}) \quad (6.11)$$

and let

$$C = \begin{bmatrix} c_1 \\ c_2 \\ \dots \\ c_m \end{bmatrix} \quad (6.12)$$

From (6.10), (6.11), and (6.12)

$$p_b = v_b + \mu_b C \quad (6.13)$$

From (5.14), (5.15), and (6.13)

$$f_b - \mu_b C = \epsilon_b \quad (6.14)$$

We can now write (6.14) to include all persons, as in (6.2) and (6.3).

$${}_b^f - {}_b^\mu C = {}_b^\epsilon \quad (6.15)$$

Equation (6.15) implies $n \times n_b$ observations from which to solve for the m parameters in C . A least squares solution is given by

$$C = ({}_b^\mu {}_b^\mu)^{-1} {}_b^\mu {}_b^f \quad (6.16)$$

However, again it might be better to introduce the constraints in the preceding models. If we let

$${}_b^{\tilde{f}} = {}_b^\mu C \quad (6.17)$$

the constraint would be

$${}_b^{\tilde{f}'} {}_b^{\tilde{f}} = {}_b^{f'} {}_b^f \quad (6.18)$$

Then the solution for C would be

$$C = ({}_b^\mu {}_b^\mu)^{-1} {}_b^\mu {}_b^f \left(\frac{{}_b^{f'} {}_b^f}{{}_b^{f'} {}_b^\mu ({}_b^\mu {}_b^\mu)^{-1} {}_b^\mu {}_b^f} \right) \quad (6.19)$$

Once C is calculated, the p_b 's for each person could be calculated from equation (6.13).

We can now construct two matrices, as follows:

$$p \equiv \begin{bmatrix} 1^{p_b'} \\ \dots \\ h^{p_b'} \\ \dots \\ n^{p_b'} \end{bmatrix} \quad (6.20)$$

$$P \equiv \begin{bmatrix} 1^{P_b'} \\ \dots \\ h^{P_b'} \\ \dots \\ n^{P_b'} \end{bmatrix} \quad (6.21)$$

Then p is the $n \times n_1$ matrix of estimates of the proportion of time each person would perform at each level, and P has the same interpretation for the observed proportions. The expected and actual performance vectors, respectively, would be given by

$$\tilde{Y} = p L_b \quad (6.22)$$

$$Y = P L_b \quad (6.23)$$

The correlation between \tilde{Y} and Y could be taken as a test of the model.

The question may well be raised as to whether the assumption of a functional relationship between the valence of a given level of effort and the expectancy of exerting that level of effort is debatable. In general, however, it does not seem unreasonable to suppose that the greater the valence of an effort level the greater would be the effort exerted at that level.

In any case, if it can be established that for some value of m (the degree of the polynomial) the C vector is adequately stable, this vector of weights can be used on new samples to estimate performance on the basis of their matrices of expectancies, P_{ab} , and their vectors of performance level valences, V_b .

APPENDIX J

A PROOF OF THE VALIDITY OF A NEBEKER EXPECTANCY THEORY MODEL

APPENDIX J

A PROOF OF THE VALIDITY OF A NEBEKER EXPECTANCY THEORY MODEL

Suppose that for a given task we have a scale of performance divided into a number of class intervals or levels. Over a given range of time, the percentage of this time that a person performs at each level is recorded. His average performance can then be calculated as the sum of the products of corresponding proportions and performance levels. Suppose also that we have some experimentally independent procedure for estimating the proportion of time that the person will perform at each level of the performance scale. Using such a set of estimated proportions, we can also estimate the average performance of the person.

If we have such data available for a number of persons, we can calculate the correlation between the observed and estimated average performance level for the group of persons. We can also investigate the extent to which the estimated percentages of time for each level for all persons correspond to the observed percentages. Under certain assumptions it can be shown that the correlation between the estimated and observed average performance is equal to the average and observed proportions of time for corresponding performance levels. These findings support a speculation by Nebeker* that the relationship between the two correlations should be high. The proof of this statement evolves from the following development.

Let

P be an $n \times m$ matrix of probabilities such that

$$P \mathbf{1} = \mathbf{1} \quad (1)$$

V be an m th order vector of an interval scale

* Personal communication.

$$Y = P V \quad (2)$$

P_a and P_b be independent estimates of P where

$$P_a' 1 = P_b' 1 = 1 \quad (3)$$

$$Y_a = P_a V \quad (4)$$

$$Y_b = P_b V \quad (5)$$

r_{ab} is the correlation between Y_a and Y_b .

It is well known that

$$r_{ab} = \frac{Y_a' (I - \frac{1 1'}{n}) Y_b}{\sqrt{Y_a' (I - \frac{1 1'}{n}) Y_b} \sqrt{Y_b' (I - \frac{1 1'}{n}) Y_b}} \quad (6)$$

Let

$$C_{ab} = \frac{1}{n} P_a' (I - \frac{1 1'}{n}) P_b \quad (7)$$

$$C_{aa} = \frac{1}{n} P_a' (I - \frac{1 1'}{n}) P_a \quad (8)$$

$$C_{bb} = \frac{1}{n} P_b' (I - \frac{1 1'}{n}) P_b \quad (9)$$

D_{ab} be a diagonal of C_{ab}

D_a^2 be the diagonal of C_{aa}

D_b^2 be the diagonal of C_{bb}

By definition, the C 's are covariance matrices corresponding to the P_a and P_b matrices and the diagonal elements of the D^2 matrices are variances. We let

R_{ab} be the matrix of correlations for the columns of P_a with those of P_b

R_{aa} be the matrix of correlations among the columns of P_a

R_{bb} be the matrix of correlations among the columns of P_b .

According to the definitions

$$C_{ab} = D_a R_{ab} D_b \quad (10)$$

$$C_{aa} = D_a R_{aa} D_a \quad (11)$$

$$C_{bb} = D_b R_{bb} D_b \quad (12)$$

From (3) through (5) and (7) through (9) it can be shown that (6) can be written

$$r_{ab} = \frac{V' D_a R_{ab} D_b V}{\sqrt{V' D_a R_{aa} D_a V} \sqrt{V' D_b R_{bb} D_b V}} \quad (13)$$

From (3) and (7) through (9) it can be proved that

$$1' C_{ab} 1 = 0 \quad (14)$$

$$1' C_{aa} 1 = 0 \quad (15)$$

$$1' C_{bb} 1 = 0 \quad (16)$$

We let

$$v = I - \frac{1 1'}{m} \quad (17)$$

From (3) and (10) through (17) we have

$$r_{ab} = \frac{v' D_a R_{ab} D_b v}{\sqrt{v' D_a R_{aa} D_a v} \sqrt{v' D_b R_{bb} D_b v}} \quad (18)$$

From (10) through (12) and (14) through (16)

$$1' D_a R_{ab} D_b 1 = 0 \quad (19)$$

$$1' D_a R_{aa} D_a 1 = 0 \quad (20)$$

$$1' D_b R_{bb} D_b 1 = 0 \quad (21)$$

Assume now that

$$D_a = J_a I \quad (22)$$

$$D_b = \sigma_b I \quad (23)$$

Then (18) becomes

$$r_{ab} = \frac{v' R_{ab} v}{\sqrt{v' R_{aa} v v' R_{bb} v}} \quad (24)$$

From (19) through (21) and (22) and (23)

$$l' R_{ab} l = 0 \quad (25)$$

$$l' R_{aa} l = 0 \quad (26)$$

$$l' R_{bb} l = 0 \quad (27)$$

Assume next that

$$R_{a_i b_i} = \delta_{ab} \text{ for all } i \quad (28)$$

$$R_{a_i b_j} = \gamma_{ab} \text{ for all } i \text{ and } j, i \neq j \quad (29)$$

$$R_{a_i a_j} = R_a \text{ for all } i \text{ and } j, i \neq j \quad (30)$$

$$R_{b_i b_j} = R_b \text{ for all } i \text{ and } j, i \neq j \quad (31)$$

Then we have from (28) through (31)

$$R_{ab} = \delta_{ab} I + \gamma_{ab}(l l' - I) \quad (32)$$

$$R_{aa} = I + R_a(l l' - I) \quad (33)$$

$$R_{bb} = I + R_b(l l' - I) \quad (34)$$

From (25) and (32)

$$\delta_{ab} m + \gamma_{ab}(m^2 - m) = 0 \quad (35)$$

From (26) and (33)

$$m + R_a(m^2 - m) = 0 \quad (36)$$

From (27) and (34)

$$m + R_b(m^2 - m) = 0 \quad (37)$$

From (35), (36), and (37) we have, respectively,

$$\gamma_{ab} = -\frac{\delta_{ab}}{m-1} \quad (38)$$

$$R_a = -\frac{1}{m-1} \quad (39)$$

$$R_b = -\frac{1}{m-1} \quad (40)$$

From (32), (33), and (34) in (38), (39), and (40), respectively, we have

$$R_{ab} = \delta_{ab} \left(I - \frac{1}{m} \frac{1}{1'} \right) \frac{m}{m-1} \quad (41)$$

$$R_{aa} = \left(I - \frac{1}{m} \frac{1}{1'} \right) \frac{m}{m-1} \quad (42)$$

$$R_{bb} = \left(I - \frac{1}{m} \frac{1}{1'} \right) \frac{m}{m-1} \quad (43)$$

From (41), (42), and (43) in (24) we have, because of (17)

$$r_{ab} = \delta_{ab} \quad (44)$$

Thus, with the assumptions (22), (23), and (28) through (31), we have:

The correlation between estimated and observed average performance is equal to each of the correlations of the estimated and observed proportions of time for corresponding performance levels.

DISTRIBUTION LIST

Military Assistant for Training and Personnel Technology (ODUSD(R&AT))
Chief of Naval Operations (OP-115) (2), (OP-987H)
Chief of Naval Material (NMAT 05), (NMAT 052), (NMA1 0722)
Chief of Naval Research (Code 200), (Code 440) (3), (Code 442), (Code 448)
Chief of Naval Education and Training (N-5)
Commander Naval Military Personnel Command (NMPC 013C)
Commanding Officer, Naval Regional Medical Center, Portsmouth, VA (ATTN: Medical Library)
Commanding Officer, Naval Training Equipment Center (Technical Library)
President, Naval War College (Code E114)
Superintendent, Naval Postgraduate School
Commander, Army Research Institute for the Behavioral and Social Sciences, Alexandria (PERI-ASL)
Chief, Army Research Institute Field Unit--USAREUR (Library)
Chief, Army Research Institute Field Unit, Fort Harrison
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base (Manpower and Personnel Division)
Commander, Air Force Human Resources Laboratory, Brooks Air Force Base (Scientific and Technical Information Office)
Commander, Air Force Human Resources Laboratory, Williams Air Force Base (AFHRL/OT)
Commander, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base (AFHRL/LR)
Program Manager, Life Sciences Directorate, Bolling Air Force Base
Commanding Officer, U.S. Coast Guard Research and Development Center, Avery Point
Superintendent, U.S. Coast Guard Academy
Director, Science and Technology, Library of Congress
Defense Technical Information Center (DDA) (12)